Using CRISPR Knockout Screens to Investigate Biological Phenomena

by

William A. Walker

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
(Chemistry)
in The University of Michigan
2020

Doctoral Committee:

Professor Ryan Bailey, Co-Chair
Assistant Professor Brent R. Martin, Co-Chair
Assistant Professor Kristin Koutmou
Assistant Professor Markos Koumous
Assistant Professor Paul Jenkins

William A. Walker
wawalke@umich.edu
ORCID iD: 0000-0002-2514-6327

TABLE OF CONTENTS

LIST OF FIGURES …….. vi

LIST OF SCHEMES …….. viii

LIST OF APPENDICES …….. ix

LIST OF ABBREVIATIONS ……... x

ABSTRACT ……..... xii

Chapter 1 Genomic Editing and Screening: Past, Present, and Future.....1

1.1 Early Genome Editing and Its Limitations.....1

1.2 The Emergence of CRISPR-Cas9 Gene Editing.....5

1.3 Cpf1: The Next Generation of CRISPR Gene Editing.....6

1.4 Summary and Outlook.....8

Chapter 2 Investigating Cellular Oxidative Stress using CRISPR-Cas9 Genomic Screens.10

2.1 Introduction …….10

2.1.1 Oxidative Stress and Cellular Life.....……10

2.2 Research Justification and Objectives ……12

2.3 Methods ……..……...……..13

2.3.1 Cell Culture ..…...……...…...……..13

2.3.2 TKO Library Amplification ……...……....14

2.3.3 Viral Production and Quantification ……...……..……...14

2.3.4 Generation of Cas9-Expressing Cell Lines..……....…….....15

2.3.5 Western Blotting ……..……....……..…....15

2.3.6 H_2O_2 Viability Assay.....16

2.3.7 TKO CRISPR-Cas9 Essential Gene and Oxidative Stress Screens.....16

2.3.8 Extraction of Genomic DNA …….…….....17

2.3.9 Illumina Library Preparation and Sequencing …….…....18

2.3.10 Data Analysis …….……..……19

2.4 Results …….....20

2.5 Discussion ……....26

2.6 Conclusions.....36

Chapter 3 Uncovering Functional Redundancy in DHHC-family Palmitoyltransferases using CRISPR-Cpf1 Duplex Genomic Screens ...…….37

3.1 Introduction …….....37

3.1.1 Protein Lipidation in Cell Biology …….……..…..……....37

3.1.2 The DHHC Family Acyltransferases ……....…….……....38

3.1.3 Inter- and Intracellular Distribution of DHHC-Family Acyltransferases...39

3.1.4 Substrate Preference and Functional Redundancy within DHHC-Family Acyltransferases ……....…….....41

3.1.5 Ras GTPases in Normal and Aberrant Cellular Processes …….....……...44

3.2 Research Justifications and Objectives ……...45

3.3 Methodology ……...47

3.3.1 Cell Culture ……..... 47

3.3.2 Design and Synthesis of mDHHC Arrays ……. 48

3.3.3 Design and Synthesis of hDHHC Arrays ……..... 48

3.3.4 Cloning of mDHHC/hDHHC sgRNA Libraries …….…… 49

3.3.5 Cpf1 Library Amplification …….……51

3.3.6 Viral production and quantification …….……51

3.3.7 Generation of Cpf1-Expressing Cell Lines …….……52

3.3.8 Western Blotting …….…….52

3.3.9 mDHHC CRISPR-Cpf1 Duplex Knockout Screens for Ras-dependent Growth ……..……..…....53

3.3.10 Extraction of Genomic DNA …… 53

3.3.11 Illumina Library Preparation and Sequencing ……..... 53

3.3.12 Data Analysis …… 55

3.4 Results …… 56

3.4.1 Cloning of hDHHC Cpf1 Library …….....……... 56

3.4.2 Cloning of mDHHC Cpf1 Library.....57

3.4.3 HAP1 Screen for DHHC Functional Redundancy ……. 59

3.5 Discussion 66

3.6 Conclusions ……... 70

Chapter 4 Conclusions and Future Directions ……... 72

4.1 Overview of Experimental Outcomes ……... 72

4.2 Considerations for Optimal crRNA Design …….. 73

4.3 Considerations for Optimal CRISPR Screen Design …… 74

4.4 General Guidelines for CRISPR Genomic Screen Design ……..……...75

4.5 Conclusions …….. 77

References 78

Appendices 90

LIST OF FIGURES

Figure 1. Early genome editing using triple-helical DNA.....2

Figure 2. Targeted genome editing using ZNFs and TALENs 3

Figure 3. Mechanism of gene suppression by RNA silencing 4

Figure 4. Mechanism of action of CRISPR-Cas9…….. 6

Figure 5. Mechanism of action of CRISPR-Cpf1.....8

Figure 6. Sequential oxidation of reactive cysteine thiols……...…….12

Figure 7. Average Log_2 Fold Changes for Essential Genes with High Representation……....21

Figure 8. Previously identified essential genes identified as hits in HAP1 control screen……...23

Figure 9. Representation of essential genes in top-scoring hits identified in control screen 24

Figure 10. Identified essential genes candidates implicated in cell survival..…….……..28

Figure 11. Identified essential gene candidates implicated in cell growth and proliferation...….29

Figure 12. Identified essential gene candidates implicated in mitochondrial maintenance….....30

Figure 13. Identified essential gene candidates implicated in lipid biogenesis……...…….30

Figure 14. Identified essential gene candidates implicated in DNA damage response…..…..….31

Figure 15. Identified essential gene candidates implicated in the cellular oxidative stress response……...……...……...32

Figure 16. Identified essential gene candidates with uncharacterized functions……..……....34

Figure 17. The various roles of DHHC-family palmitoyltransferases in normal cellular function……...……...38

Figure 18. Sequence homology relationships between DHHC-family palmitoyltransferases.....44

Figure 19. Representation of crRNA arrays in hDHHC Cpf1 Library…… 56

Figure 20. Representation of all crRNAs by count in human hDHHC Cpf1 library……..... 57

Figure 21. Representation of crRNA arrays in mDHHC Cpf1 Library……..... 58

Figure 22. Representation of all crRNAs by count in mDHHC Cpf1 library……... 58

Figure 23. Comparison of average library representation in the hDHHC and mDHHC CRISPR Cpf1 libraries……...……..... 59

Figure 24. Volcano plots of all recovered crRNA arrays from HAP1 hDHHC library screens...60

Figure 25. Average $\log_2$ fold change of single targeting crRNA arrays exhibiting depletion.....61

Figure 26. Average $\log_2$ fold change of all single targeting crRNA arrays exhibiting enrichment. 62

Figure 27. Average $\log_2$ fold change of all single targeting crRNA arrays compared to duplex arrays for all DHHC structural subfamilies…….. 63

Figure 28. Volcano plots of nontargeting and essential gene arrays.....65

Figure 29. Average $\log_2$ fold change of crRNA arrays for previously validated essential genes 65

Figure 30. An alternative approach for designing CRISPR-Cpf1 knockout libraries.....69

Figure 31. Acyl-RAC enrichment for LC-MS proteomics.....70

LIST OF SCHEMES

Scheme 1 Diagrammatic representation of methods used for screening of oxidative stress factors by CRISPR knockout screening…….13

Scheme 2 Methods to be used for Cpf1 library synthesis and subsequent pooled KO screen…. 47

LIST OF APPENDICES

Appendix A. Cloning, Expression and Validation of pCLHCX-AsCpf1…… 90

Appendix B. PCR Primers for Illumina Library Preparation……... 93

Appendix C mDHHC CRISPR-Cpf1 Library crRNA Sequences……..100

Appendix D hDHHC CRISPR-Cpf1 Library crRNA Sequences……...104

LIST OF ABBREVIATIONS

Acyl-RAC	Acyl resin-assisted capture
CRISPR	Clustered Regularly Interspersed Short Palindromic Repeats
CRISPRa	CRISPR activation
CRISPRi	CRISPR inhibition
crRNA	CRISPR RNA
DKO	Double Knockout crRNA
ER	Endoplasmic Reticulum
GA	Golgi Apparatus
GAP	GTPase-Activating Protein
gDNA	Genomic DNA
GDP	Guanosine Diphosphate
GEF	Guanine-Nucleotide-Exchange Factor
GTP	Guanosine Triphosphate
MOI	Multiplicity of Infection
NHEJ	Non-homologous End Joining
NT	Non-targeting crRNA
PAM	Protospacer Adjacent Motif
PAT	Protein Acyl Transferase
PTMs	Post-translational Modifications

PM	Plasma Membrane
ROS	Reactive Oxygen Species
RPM	Reads per Million
sgRNA	Single guide RNA
shRNA	Small Hairpin RNA
siRNA	Small interfering RNA
SKO	Single Knockout crRNA
TALEN	Transcriptional activator like effector nuclease
TKO	Toronto Knockout Library
tracrRNA	Trans-activating CRISPR RNA
ZNF	Zinc Finger Nuclease

Abstract

The discovery of CRISPR has revolutionized the study of life by providing tools for the precise and controlled manipulation of an organisms' genome. Knocking out, suppressing, enhancing, or replacing select genes using CRISPR have revealed their unique contributions to an organisms' survival with an ease that was unachievable with the prior state-of-the-art methods. This innovation has extended to whole-genome screening, previously conducted with small hairpin RNA (shRNA) libraries. CRISPR-Cas9 pooled library screens have been shown to effectively identify genes essential to a number of physiological processes including drug toxicity and viral infection. Herein, we used previously designed CRISPR-Cas9 knockout libraries to perform whole genome screens on cells subjected to oxidative stress to discovery previously unidentified genes that provide protection under these conditions. Conducting these screens revealed previously identified essential genes in the untreated population in accordance with our expectations. Performing the screens in cell populations under oxidative stress revealed several individual genes with putative roles associated with or accessory to stress tolerance but failed to yield a uniform collection of genes clearly identifiable with this physiological response.

Recent efforts to improve upon existing CRISPR methods have centered around using the related Cpf1 nuclease. Cpf1 offers several advantages over Cas9, the most important being its ability to process multiple targeting RNAs from a single expression vector and enabling the simultaneous knockout of multiple genes. With this in mind, we designed randomized duplex

Cpf1 libraries targeted to the DHHC-family palmitoyltransferases (PAT). The DHHC PATs constitute a large family of structurally related protein acyltransferases for which substrates and the rules that dictate their specification are still largely unknown. These Cpf1 libraries were then used in growth screens with the hope of determining to what extent, if any, that functional redundancy existed in the substrate profiles of DHHC PATs. Initial screens using these libraries revealed no identifiable change in cell growth for any combination of DHHCs. This was also the case for the included essential gene controls, indicating that inefficiency in Cpf1 activity was to blame for the apparent lack of discernable hits generated by the screen. This result was supported by recent literature that found that unmodified Cpf1 significantly underperformed Cas9 when used to perform pooled screens.

Although our work with Cas9 and Cpf1 CRISPR screens did not yield the results desired at the outset of this work they did provide us with several considerations that may serve as useful guidelines for others who wish to make use of these powerful genetic tools.

Chapter 1

Introduction

Genomic Editing and Screening: Past, Present, and Future

1.1 Early Genome Editing and Its Limitations

The discovery of the DNA double helix led in short order to a search for methods to manipulate its sequence. The discovery of DNA repair and recombination provided inspiration for the first successful demonstrations of selective editing of the DNA sequence.[1] These made use of complementary sequences between the targeted region of the genome and an exogenous donor plasmid to drive homologous recombination between them and stably introduce selective modifications to the genome.[2,3] The efficacy of these methods was later improved upon by introducing precise breaks in the DNA at desired locations within the genome using chemical or cross-linking agents.[4,5](**Figure 1**)

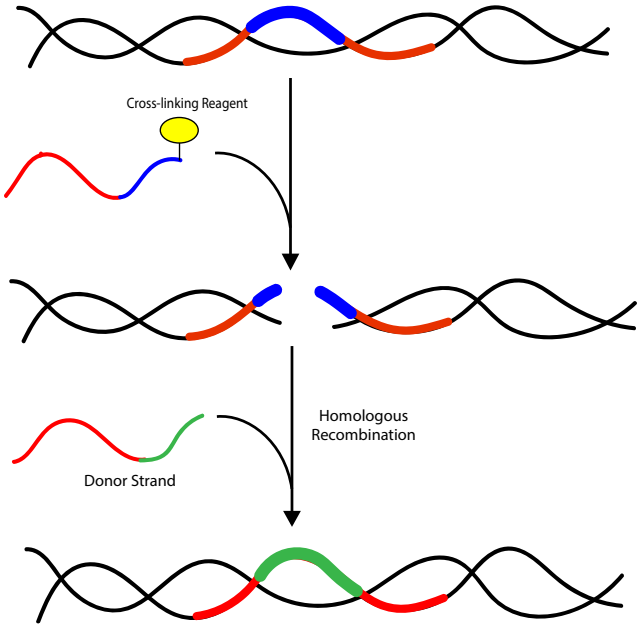


Figure 1. Early targeted genome editing made possible thorough the introduction of synthetic template oligonucleotides was improved by the concurrent introduction of complementary triple-helical DNA modified with a crosslinking agent to induce double stranded breaks in the genome.

However, even with these improvements the prevalence of the desired genomic modifications hardly ever approached 1.0 percent of the subject population, rendering them impracticable for anything more than the simplest investigations.[6,7] The discovery of first zinc finger nucleases (ZNFs) and then transcriptional activator-like effector nucleases (TALENs) offered the possibility of improvement over these early methods by providing greater specificity and flexibility in the selection of genomic targets.[8] These methods combined the DNA-binding domains of zinc finger and transcriptional activator-like effector proteins with the nonspecific nuclease domain of the FokI restriction endonuclease to produce a synthetic targeted endonuclease capable of editing specific regions of the genome.[9,10] A feature critical to the specificity of these nucleases is the requirement that the FokI domain dimerize to enable cutting of the DNA, necessitating the design of two unique DNA binding domains for each genomic target.[11](**Figure 2)**

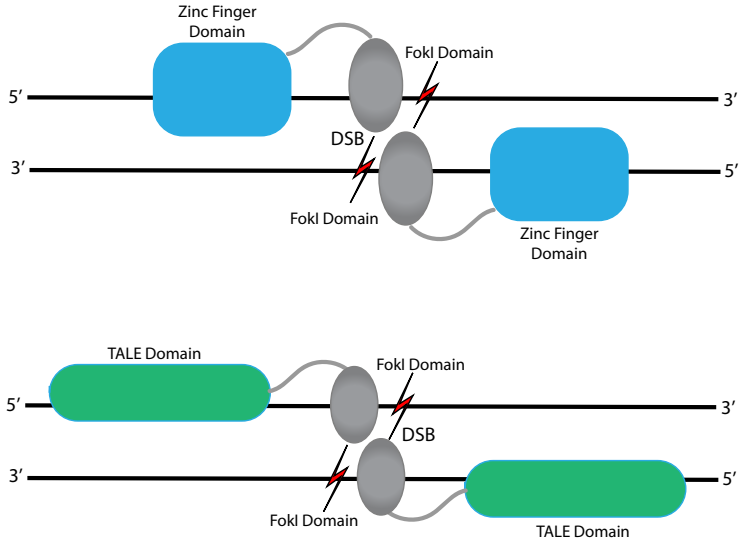


Figure 2. ZNFs (top) and TALENs (bottom) are engineered fusions of zinc finger and transcription activator like effector DNA binding and FokI nuclease domains. Pairs of ZNFs and TALENs are designed to recognize specific regions of the genome and induce double-stranded breaks upon dimerization of the FokI domains.

This has been shown to permit controlled editing of the genome with minimal risk of off-target effects and demonstrably improved editing efficiency over earlier methods.[12] However, in spite of these advantages ZNFs and TALENs have only seen limited use due to the inherent difficulties of designing and validating heterodimeric pairs for every DNA locus of interest.[13] RNA-mediated gene silencing provides an alternative approach to addressing the challenge of controlling gene expression. Whether introduced directly as double-stranded small interfering RNAs (siRNAs) or through expression vectors as small hairpin RNAs (shRNAs) these small RNAs operate to control gene expression through fundamentally different mechanisms to those that operate at the level of the genome and, in doing so, address many of their inherent shortcomings. RNA silencing suppresses gene expression by targeting and degrading mRNA transcribed from the genome rather than the genome itself, making its effects reversible and far more amenable to the study of genes that would otherwise be lethal if permanently depleted by

gene knockout.[14] Furthermore, because RNA silencing operates on the basis of nucleotide sequence complementarity it can be easily scaled to suppress the expression of any number of genes desired even to the level of the genome.[15](**Figure 3)**

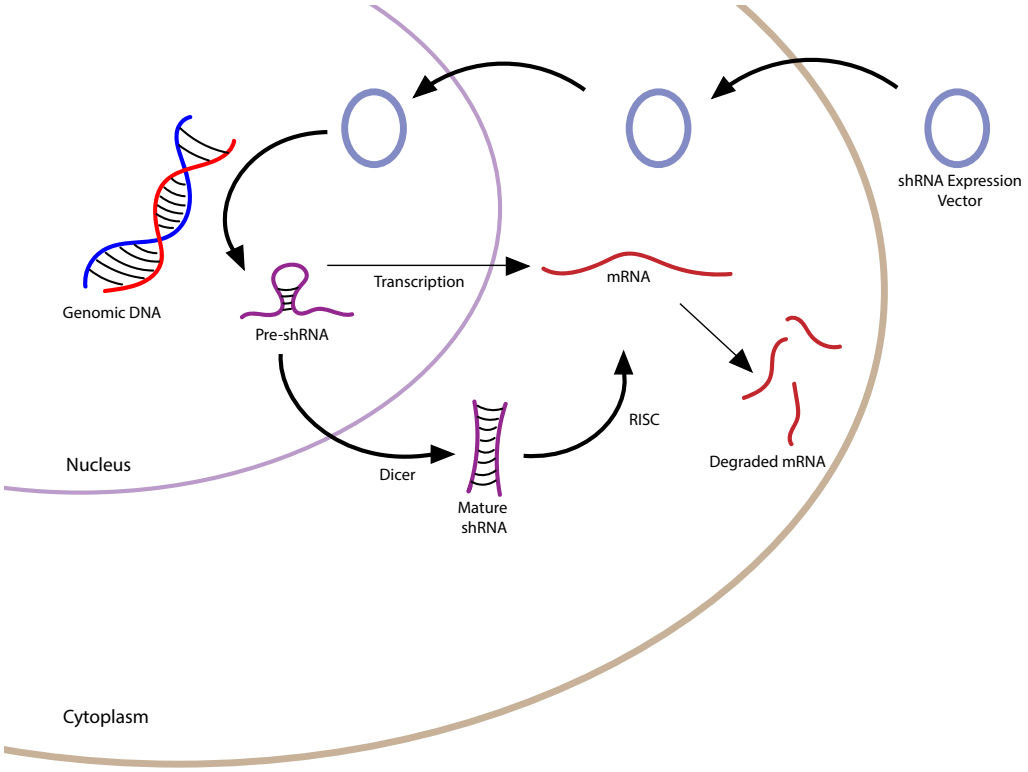


Figure 3. Targeted suppression of gene expression by the stable introduction of shRNA expression vectors.

This flexibility made possible the first global investigations of gene function using shRNA expression libraries targeting large numbers of known protein-coding genes.[16] Pooled screens performed with these libraries led to the identification of genes critical to cellular function[17,18], drug action[19], and viral pathogenesis[20] among other biological phenomena. Whatever advantages RNA silencing held over nuclease-dependent methods of gene editing in scalability it lacked in efficiency and specificity. Different sequences used to induce RNA silencing would often result in variable levels of suppression with sometimes even the minimal levels of expression being

sufficient to maintain a detectably normal phenotype.[21] Furthermore, it was hardly uncommon for sequences targeted to one particular gene to exert unintended but measurable effects on others, thereby confounding the interpretation of results.[22] To this point the genetic tools that had been developed provided a means, although imperfect, to interpret gene function although an ideal system of gene editing, one that combined the scalability and flexibility of RNA silencing with the efficiency and specificity of engineered nucleases remained out of reach.

1.2 The Emergence of CRISPR-Cas9 Gene Editing

Since its discovery in *S. thermophilus*[23] and subsequent demonstration as a gene-editing tool in species other than bacteria[24], the CRISPR-Cas9 system has been elaborated into a versatile set of tools for specific gene editing in organisms spanning all domains of life. Cas9 is a non-specific nuclease that, once combined with short guide RNAs (crRNAs), localizes to complementary regions within the genome, creates double-stranded breaks around the target sequence, and removes it from the genome.[25](**Figure 4)** Once synthesized, pre-crRNAs containing a targeting sequence interact with a complementary trans-activating RNA to form a mature guide RNA (sgRNA) that is recognized by Cas9 and then subsequently directs it to complementary regions within the genome, recognizing a 20bp sequence proximal to a short, protospacer adjacent motif (PAM) (5'-NGG-3' for the often used spCas9).[26] To repair the double-stranded break, cells generally rely on the more efficient but error-prone nonhomologous end-joining (NHEJ), resulting in inactivating mutations when the Cas9 nuclease is targeted to protein-coding regions of the genome.[27] Since Cas9 only generates inactivating mutations it avoids any issues with incomplete suppression of gene expression that plague siRNA-mediated knockdown. Furthermore, Cas9 has been shown to exhibit low off-target effects[28] and so when applied in a pooled screen format similar to that used for shRNAs provides a more unambiguous

determination of gene function at the level of the genome for any process under investigation. Traditional Cas9 knockout methods have also been modified for genome screening by transcriptional inhibition or activation of target genes using inactivated Cas9 (dCas9) activator/repressor fusions[29,30,31], providing a versatile platform for performing genome-wide screens. These methods together have been used in support or in place of shRNA-based screens to identify previously unknown essential genes[32,33,34], oncogenes[35], mediators of viral replication[36,37], and drug action[38].

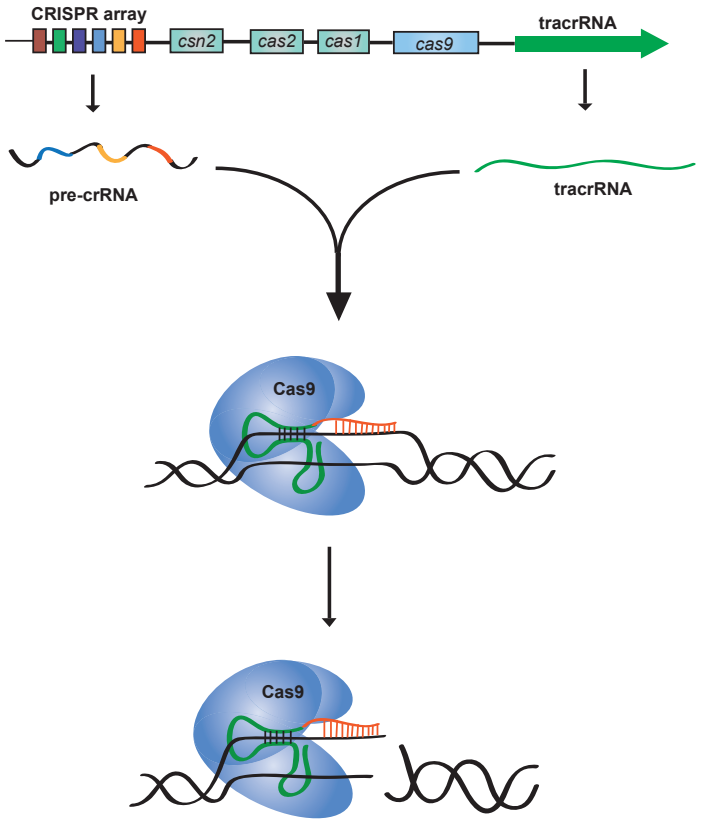


Figure 4. Mechanism of action of CRISPR-Cas9. crRNAs combine with tracrRNA in the native CRISPR system to form mature sgRNAs. These are recognized through the stem-loop structure of the tracrRNA by the Cas9 nuclease which is then directed to its genomic target by the complementary 20bp sequence of the crRNA where it creates a double stranded break in the DNA.

1.3 Cpf1: The Next Generation of CRISPR Gene Editing

A recent addition to the growing number of known CRISPR-Cas systems, Cpf1 was discovered to be a type II system similar in make to Cas9, composed of a single large Cas effector protein and relying on complementary crRNAs to mediate genetic interference.[39] There are, however, several functional differences that distinguish Cpf1 from its better known relative. Unlike Cas9 systems, Cpf1 crRNAs do not require a trans-activating crRNA (tracrRNA) to process CRISPR arrays into mature crRNAs.[40] Cpf1 also recognizes and cleaves DNA sequences preceded by a T-rich (TTTN) protospacer-adjacent motif (PAM) and generates a staggered double-stranded break in the DNA outside of the target sequence as opposed to the blunt-end cut internal to target sequence produced by Cas9.[41] These functional differences consequently lend several advantages to Cpf1 over Cas9 when used for targeted gene editing. Chief among these is the ability of Cpf1 to process its own mature crRNAs without assistance from endogenous cellular components.[42] This allows for multiple crRNAs to be expressed from a single Pol III promoter and creates opportunities for exploring multiplex gene editing and screening whereas previously this only possible by either expressing single crRNAs from multiple promoters or simultaneous use of multiple plasmids with Cas9.[43](**Figure 5**) Cpf1 has also been shown to cleave DNA distal to the target site , promoting the formation of indels there and preserving the target site for additional rounds of cleavage.[40] Moreover, the ability of Cpf1 to carry out gene editing from these arrays has been demonstrated across multiple genetic models including animals and plants.[44]The utility of these unique attributes was recently demonstrated by using Cpf1 to conduct randomized duplex genetic knockout screens, all the while leaving open the possibility of carrying out similar screens of higher multiplicity; something that was previously unachievable with Cas9.[45]

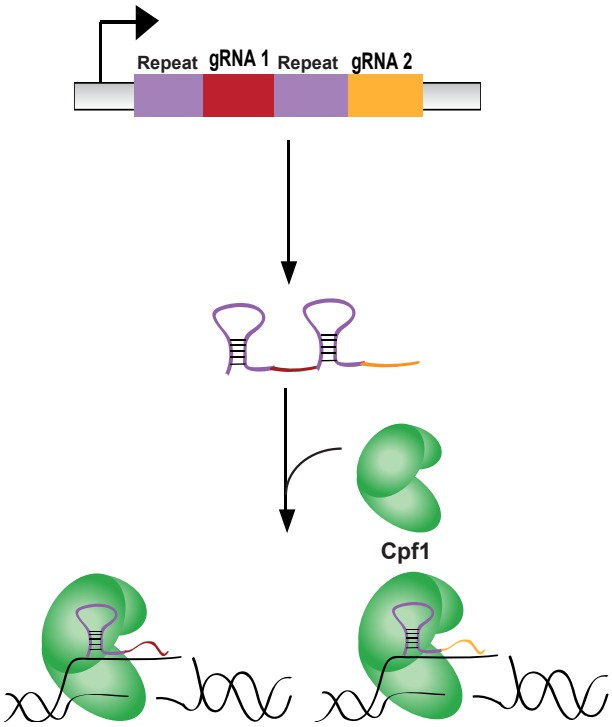


Figure 5. Processing of crRNAs by Cpf1. Unlike Cas9, Cpf1 does not require a tracrRNA to bind to its crRNAs and is capable of processing multiple crRNAs from a single continuous transcript.

1.4 Summary and Outlook

The advantages of CRISPR-based genomic screens over preceding methods have been repeatedly demonstrated since their introduction. The following chapters detail our efforts to harness the unmatched flexibility and accuracy of CRISPR genomic screens to gain greater insight into complex biological phenomena.

In **Chapter 2**, we detail our attempt the use a whole genome CRISPR-Cas9 screen to identify critical components of the oxidative stress response. Here, limitations in the design of the screen with regard to the application of genotoxic stress prevented the identification of our intended gene targets, demonstrating the importance proper design and validation in the performance of these genomic screens.

In **Chapter 3**, we go over our endeavors to design and apply a novel CRISPR-Cpf1 duplex knockout library to investigate questions of functional within a family of protein acyltransferases. Here, previously unknown functional limitations of CRISPR-Cpf1prevented us from accomplishing our stated experimental objectives but do provide valuable insight into the use of this system for future endeavors.

In **Chapter 4**, we summarize our experimental outcomes and use them to evaluate the limitations of CRISPR genomic screening. We then provide recommendations on the use of CRISPR-based methods in future work on the basis of these and recently published literature.

Chapter 2

Investigating Cellular Oxidative Stress using CRISPR-Cas9 Genomic Screens

1.1 Introduction

1.1.1 Oxidative Stress and Cellular Life

Reactive small molecules produced as the byproducts of normal metabolism are a feature of all cellular life. These reactive oxygen species (ROS) have the potential to cause severe dysfunction if left unchecked by the cell.[46,47] This led early on in the history of life to the evolution of stress responses to mitigate the deleterious effects of these metabolic byproducts.[48] The components of these stress responses can be found in some form in all of life, ranging from simple antioxidative enzymes of bacteria and archaea to the elaborate networks found in multicellular organisms.[49] The sheer genomic complexity of higher organisms complicates the identification of genes involved in these stress responses and their individual contributions to preventing cellular damage. A further impediment to understanding these genes play centers around the essential role that these reactive small molecules play in normal cellular signaling. Normal cellular function therefore requires that a balance be maintained between their beneficial aspects and the damage that they will inevitably cause if left unchecked.[50] The overall redox balance of the cell is maintained through active and passive mechanisms by an interrelated group of proteins. These include the peroxiredoxins, superoxide dismutase, and catalase which catalyze the conversion of ROS into less reactive forms, thioredoxins which are responsible for reducing cysteine residues that have been oxidized by ROS, and glutathione/glutaredoxin which function as a buffer to the

rest of the cell by scavenging free ROS.[51] Effective signaling via ROS requires a localized disruption of this imbalance while maintaining the overall redox state of the cell. These localized oxidative bursts may be caused by the environment or by endogenous sources such as NADPH-oxidase or nitric oxide synthase. The sudden increase in ROS overwhelms the localized buffering capacity of antioxidative enzymes and permits oxidation of cysteine residues.[52] Depending on the local concentration of ROS, the thiol groups of cysteine residues may be oxidized sequentially to form sulfenic, sulfinic, and sulfonic acids respectively.[53](**Figure 6**) Of these only the short-lived sulfenic and more stable sulfinic acid modifications can be reversed.[54] Oxidation of cysteine residues to any of these species has been shown to alter protein function either directly by altering the structure of the target protein or indirectly by blocking alternative post-translational modifications (PTMs) such as phosphorylation.[52,55] Such alterations may be activating or deactivating depending on the protein in question and may influence signal transduction by altering enzymatic activity, localization, and/or protein-protein interactions.[54] The unique chemistry of these cysteine oxoforms has led to the development of selective molecular probes that can be used to identify proteins containing these modifications.[56] These probes have led to the identification of a number of proteins containing oxidized cysteine residues with functions ranging from signal transduction, redox homeostasis, and metabolism to DNA repair.[57] Given the central importance of ROS to normal cellular function it may be presumed that there are yet many other proteins that play a part or are influenced by oxidative stress that have yet to be identified. The role that these proteins are known to play in pathological states such as cancer, neurodegeneration, and ageing necessitate that further efforts be made to identify all proteins that play a role in modulating oxidative stress.

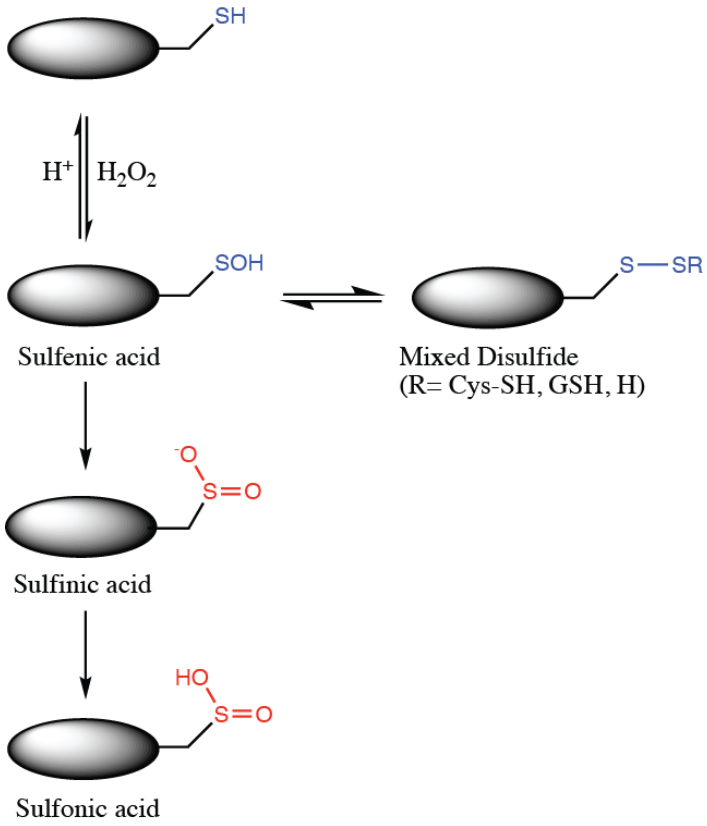
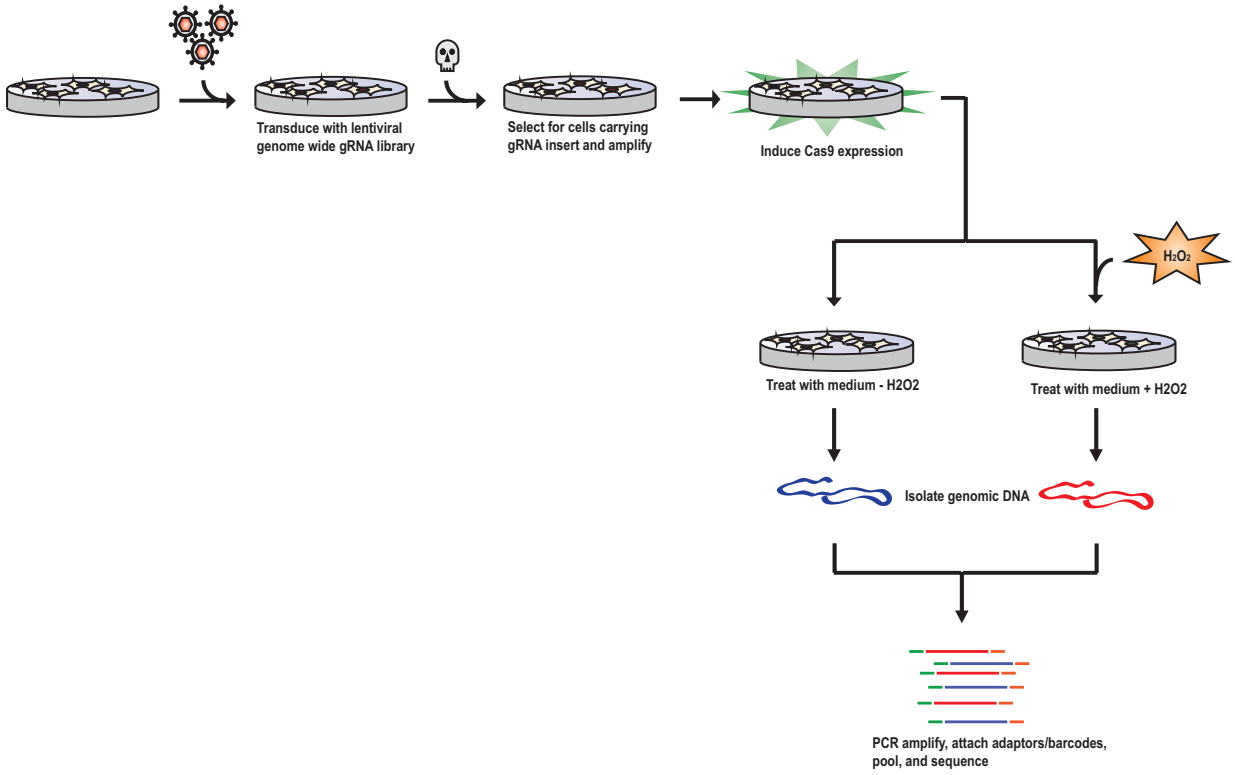


Figure 6. Sequential oxidation of cysteine by reactive oxygen species modulates protein activity.

1.2 Research Justification and Objectives

Previous efforts to identify the components of the cellular oxidative stress response using small molecule probes have revealed a large number of proteins spanning a range of functional classes with many as yet to be identified. To date, only limited efforts have been made to fully characterize the components of the cellular oxidative stress response using modern genomic screening methods. Previous attempts to accomplish this relied on sub-optimal screening methods using libraries that provided only partial genomic coverage[58,59]. We therefore hypothesize that there exist numerous genes that have yet to be implicated in the cellular oxidative stress response. CRISPR-Cas9 pooled knockout libraries currently represent the state-of-the-art in genomic screening. Using a validated whole-genome CRISPR library (Toronto

Knockout Library[33]) we will attempt to identify the genes involved in maintaining cellular redox homeostasis. (**Scheme 1**) We anticipate finding multiple genes, both characterized and uncharacterized, that have never been implicated in oxidative stress. By identifying these proteins and determining their roles in redox homeostasis we hope to arrive at a new understanding of oxidative stress and the factors that lead to abnormal function and the creation of disease states such as those seen in cancer and neurodegeneration.



Scheme 1. Diagrammatic representation of methods used for screening of oxidative stress factors by CRISPR knockout screening.

1.3 Methodology

1.3.1 Cell Culture

HAP1 cells were obtained from the laboratory of G. Van der Goot (Ecole Polytechnique Fédérale de Lausanne) and cultured in Iscove's Modified Dulbecco's Medium (Gibco) supplemented with 10% v/v Fetal Bovine Serum (HyClone) and 1% v/v penicillin/streptomycin

(Gibco). HEK293T cell lines were maintained Dulbecco's Modified Eagle's Medium (Gibco) supplemented with 10% v/v Fetal Bovine Serum (HyClone) and 1% v/v penicillin/streptomycin (Gibco).

1.3.2 TKO Library Amplification

The Toronto human knockout pooled library (TKO) was obtained as a gift from Jason Moffat (Addgene #1000000069). For each sub-library (Base 90k library, 91,320 gRNAs; Supplemental library 85,180 gRNAs) 1µl of plasmid DNA was added in separate transformations to Endura Electrocompetent Cells (Lucigen) and electroporated according to manufacturer's instructions. After 1hr. of incubation in recovery medium (Lucigen), 10µl of culture was removed, serially diluted 4x, and plated on ampicillin agar to assess transformation efficiency. The remaining transformed cells were added to 100ml LB medium containing 100µg/ml ampicillin and cultured overnight at 30°C. Plasmid DNA for each sub-library was then harvested from their respective cultures using a Plasmid Plus Maxi Kit (Qiagen) according to the manufacturer's instructions.

1.3.3 Viral production and quantification

HEK293T cells were plated at 30-40% confluence in 15cm Biolite culture dishes (Thermo Scientific) and incubated for 24hrs prior to transfecting with TKO library plasmid (16µg; 1:1 Base:Supplemental Libraries), psPAX2 (16µg) and PMDG.2 (16µg) with 144µl Fugene HD (Roche) in Opti-MEM (Gibco). The cells were incubated for 24hrs. before aspirating and replacing with fresh DMEM culture medium. The medium was then harvested at 48 and 72hrs. post-transfection, sterile-filtered, and stored as l mL aliquots. Viral titer was determined by plating HAP1 cells at 5000 cells/well on 96-well culture plates containing 10-fold serially diluted

TKO viral medium. After 24hrs., the culture medium was replaced with fresh IMDM medium containing 1µg/ml puromycin and cultured another 24hrs. Cell viability was then assessed using a WST-1 cell proliferation assay (Takara) and the IC_{50} determined to calculate viral titer.

1.3.4 Generation of Cas9 Expressing Cell Lines

HEK293T cells were passaged and plated at 30-40% confluence on 6cm Biolite culture dishes(Thermo Scientific) and incubated for 24hrs. prior to transfecting with lenticas9-Blast (Addgene #52962) (833ng), psPAX2 (833ng), and PMDG.2 (833ng) with 7.5µl of Fugene HD transfection reagent (Roche) according to the manufacturer's instructions. The cells were incubated for 24hrs. before aspirating and replacing the medium and incubating for an additional 24hrs. The medium was then harvested at 48 and 72hrs. following transfection and stored at -80°C. Cas9-expressing HAP1 cell lines were derived from parent cultures by reverse transduction with Cas9 lentivirus and selecting for positive clones by incubation with IMDM medium containing 10µg/ml Blastocidin for 72hrs. Expression of Cas9 was then confirmed by western blot of whole-cell lysates.

1.3.5 Western Blotting

Cells were pelleted by centrifugation, resuspended in 1x DPBS (Gibco) and lysed by sonication. Total protein content was quantified by Lowry assay (Biorad) and 50µg of total protein diluted to final volume of 25µl with DPBS. 1x Lamelli Buffer + SDS was added to the lysates and boiled for 15min at 95°C. Lysates were then vortexed, centrifuged, run on 7% SDS tris-acetate gels, and transferred at 40mV overnight. Transfer membranes were blocked for 1hr with Odyssey Blocking Buffer (Licor) and rocked overnight in OBB + Cas9 (7A9-3A3) Mouse mAb (1:1000)

(Cell Signaling Technology). The blots were then washed with 1xTBST and incubated in anti-mouse stable peroxidase-conjugated goat antibody (Invitrogen) for 1hr. at room temperature. The blots were then washed again, treated with SuperSignal West Pico PLUS Chemiluminescent substrate (Thermo Scientific) and imaged with an Azure c600 gel imager (Azure Biosystems).

1.3.6 H_2O_2 Viability Assay

Cells were then plated at 5,000 cells/well with 6x replicates on 96-well culture plates and incubated for 24hrs. before treating with 0-200μM H_2O_2 in DPBS over two consecutive days with fresh culture medium being added to all wells prior to each treatment. On the day following the final treatment with H_2O_2 the culture medium from all wells was aspirated, the cells washed with DPBS, and 80μL of DPBS added to each well. To these was added 20μL of Resorufin-AM (25mM in DPBS) and the cultures incubated for 1hr. before reading the fluorescence of the activated dye at 560nm using a Tecan Infinite F500 microplate reader.

1.3.7 TKO CRISPR-Cas9 Essential Gene and Oxidative Stress Screens

Cas9-expressing HAP1 cultures were expanded and transduced with TKO lentivirus at a MOI of ≈0.3. After 24hrs incubation, the cell medium was aspirated and replaced with medium containing 1μg/ml puromycin and selection allowed to proceed over 72hrs. All selected cells were then trypsinized, pooled, and plated at 2.25 x 10^7 cells (125-fold coverage) in 15cm culture dishes (30 control: 30 treatment, total coverage 3,800 per condition) and the pooled population was sampled at 4 x 10^7 as a screen base control and denoted T_0. The cultures were incubated 24hrs to expand to ≈200-fold coverage before treating with 100μM H_2O_2 solution (treatment) in DPBS or DPBS (control) every 24hrs over two days. 72 hrs. after initiating treatment, control

and treatment populations were pooled, sampled, replated at 3.75 x 10^7 cells/plate, and incubated a further 24hrs prior to initiating the next cycle of treatment. This process was repeated over four additional cycles (denoted T_7, T_{10}, T_{13}, T_{16}) with increasing concentrations of H_2O_2 ramped in 50µM increments (150, 200, 250, 300µM respectively) with the populations sampled at the aforementioned density at the completion of each cycle. Cell counts were performed on the pooled population at the end of each cycle to ensure equal coverage between populations. All sampled time points were stored at -80°C until the completion of the screen.

1.3.8 Extraction of Genomic DNA

Cell pellets were thawed and resuspended in 500µl of 1 x TE. To the resuspended cells was then added 500µl of 2x lysis buffer (20mM Tris pH 7.5, 10mM EDTA pH 8.0, 100mM NaCl, 1% SDS) and 50µl proteinase K solution (Qiagen). The lysates were incubated at 52°C for 1hr. to ensure complete lysis. The lysates were then split into equal 500µl volumes and diluted with 500µl 1x TE before adding 1mL Phenol:Chloroform:Amyl alcohol (50:49:1) (Sigma). The emulsions were mixed by inversion for 5min. before centrifuging at 11,000 rpm for 5min. The aqueous phase was then transferred to fresh microcentrifuge tubes and an additional 1mL of Phenol:Chloroform:Amyl alcohol added. The solutions were mixed by inversion for 5min. before centrifuging for 5min. at 11,000rpm. The aqueous layer was then transferred to new microcentrifuge tubes to which was then added 100µL of NaOAc (3M; pH 5.2) and two volume equivalents of 100% EtOH. The solutions were mixed by inversion and refrigerated at -20°C to precipitate the genomic DNA. The solutions were centrifuged at 11,000 rpm for 5min. to pellet the DNA. The EtOH was removed, the pellet washed with 500µL of 70% EtOH and centrifuged again at 11,000rpm. The EtOH was removed and the pellet allowed to dry for ≈5min. before

dissolving in sterile TE buffer. The total recovered gDNA was quantified using a Tecan NanoQuant plate and concentrations normalized by agarose in-gel fluorescence.

1.3.9 Illumina Library Preparation and Sequencing

sgRNAs were recovered from genomic DNA and prepared for Illumina sequencing by nested PCR using methods and primers adapted from Moffat et al.

First (outer) PCR:

Genomic DNA :	3.2µg
$MgCl_2$ (50mM) :	1µL
dNTP Mix (10mM):	1µL
Outer Primer Mix F/R (10µM):	5µL
Herculase Buffer:	10µL
Herculase II Fusion Polymerase (Agilent):	0.5µL
diH_2O:	to 50µL

Thermocycler conditions:

1. 98°C 3min.
2. 98°C 20s.
3. 52°C 30s.
4. 72°C 30s. ➔ repeat steps 2-4 for 22 cycles
5. 72°C 3min.
6. 4°C ∞

16x replicate reactions were conducted for each sample in the screen to ensure sufficient coverage. All sample replicates were pooled following completion of the first PCR.

Second (Barcode) PCR:

Pooled outer PCR product:	4µL

$MgCl_2$ (50mM) :	1µL
dNTP Mix (10mM):	1µL
i5 Barcode Primer F (10µM):	4µL
i7 Barcode Primer R (10µM):	4µL
Herculase Buffer:	10µL
Herculase II Fusion Polymerase (Agilent):	0.5µL
diH_2O:	to 50µL

Thermocycler conditions:

1. 98°C 3min.
2. 98°C 20s.
3. 52°C 30s.
4. 72°C 30s. ➔ repeat steps 2-4 for 15 cycles
5. 72°C 3min.
6. 4°C ∞

All barcode PCR reactions were conducted in duplicate, pooled upon completion and concentrated by spin column purification (Zymo). The concentrated PCR products were resolved on 2% agarose gels and ≈200bp barcoded products excised and gel purified using a QIAquick gel extraction kit (Qiagen). The prepared Illumina libraries were sequenced using an Illumina NextSeq 550 with 150-cycle, paired-end, high-output settings.

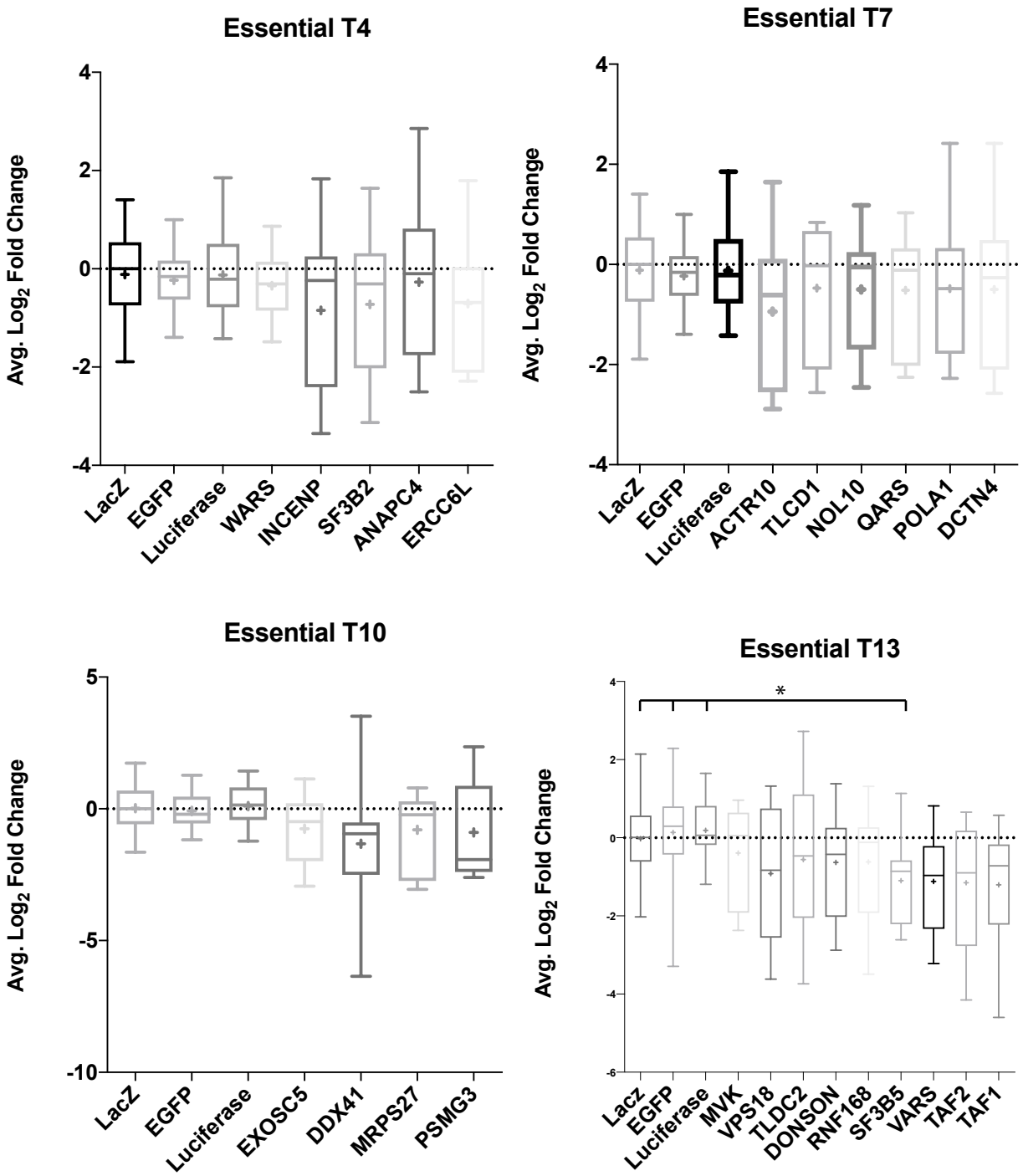
1.3.10 Data Analysis

The raw fastq data was uploaded to the Galaxy server and all subsequent processing conducted therein. In brief, forward and reverse raw fastq files were collapsed into single datasets according to their respective sample identities. The read files were then trimmed to remove 5' and 3' ends of all sequences outside of the 20bp sgRNA sequence. The quality of the reads in each file was then assessed using FASTQC to determine if quality filtering was necessary prior to downstream processing. All trimmed read files were then filtered with Trimmomatic using the

AVGQUAL 20 setting. The processed reads were then aligned to all sgRNAs present in the TKO1 Base and Supplemental Libraries using Bowtie2 using the default settings. Read counts were then extracted using SAMtools. The raw reads were then piped into the analysis program MAGeCK to using default settings to generate ranked lists of candidate genes.

1.4 Results

HAP1 cultures were transduced with the TKOv1 CRISPR KO library and treated with DPBS (control) or increasing concentrations of H_2O_2 over a period of 16 days. Following completion of the screen, populations sampled from each time point were sequenced, normalized by RPM, and fold changes calculated from the original population (T_0). All sgRNAs exhibiting a fold change of at least -2 were filtered out and those targeting core essential genes as determined by Hart et al. with a representation of at least three sgRNAs were selected to assess the performance of the screen. Average Log_2 fold changes were calculated for each essential gene and compared to the EGFP, LacZ, and Luciferase non-targeting controls. Although many of these essential genes showed decreased average fold changes over the non-targeting controls, few of these reached significant values due to the high degree of variability in the performance of individual sgRNAs targeted to the same gene. (**Figure 7**) This inefficiency has been previously noted by others and may result from a combination of factors including the nucleotide preference at the first position of the PAM, differences in chromatin structure at the target region, and cell line specific variability. Such inefficiency complicates the identification of essential genes.


Essential T4
Avg. Log2 Fold Change
4
2
0
-2
-4
LacZ
EGFP
Luciferase
WARS
INCENP
SF3B2
ANAPC4
ERCC6L
Essential T7
Avg. Log2 Fold Change
4
2
0
-2
-4
LacZ
EGFP
Luciferase
ACTR10
TLCD1
NOL10
QARS
POLA1
DCTN4
Essential T10
Avg. Log2 Fold Change
5
0
-5
-10
LacZ
EGFP
Luciferase
EXOSC5
DDX41
MRPS27
PSMG3
Essential T13
Avg. Log2 Fold Change
*
4
2
0
-2
-4
-6
LacZ
EGFP
Luciferase
MVK
VPS18
TLDC2
DONSON
RNF168
SF3B5
VARS
TAF2
TAF1

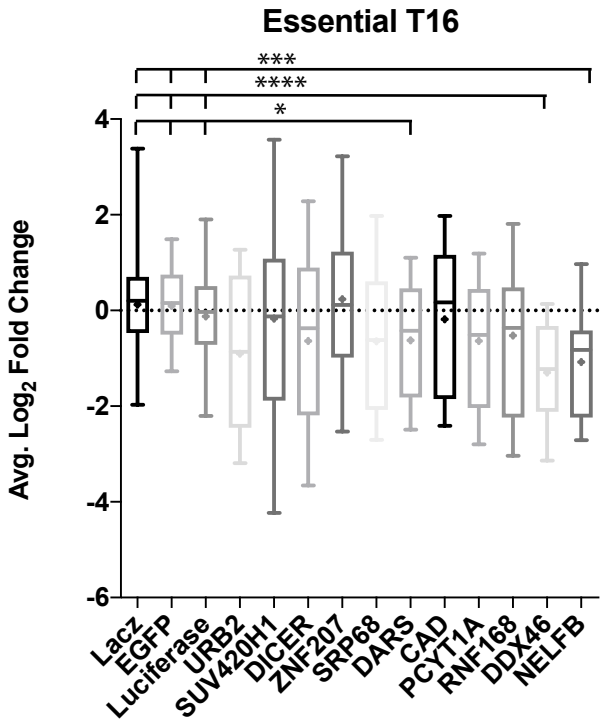
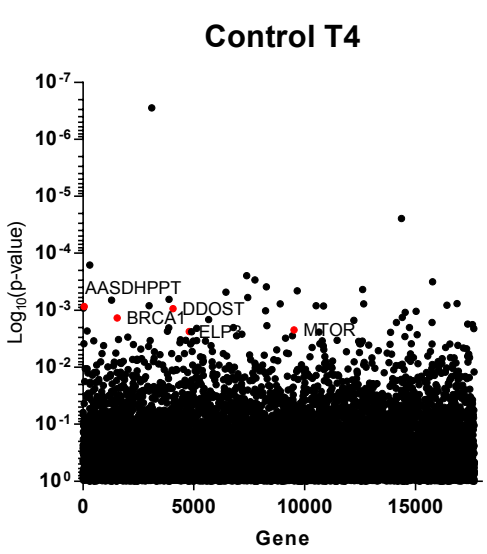
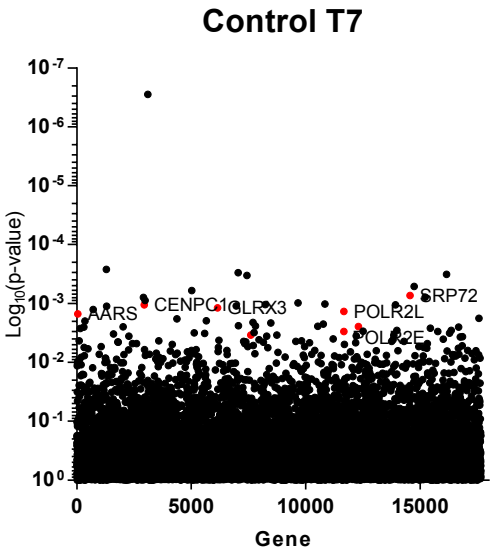
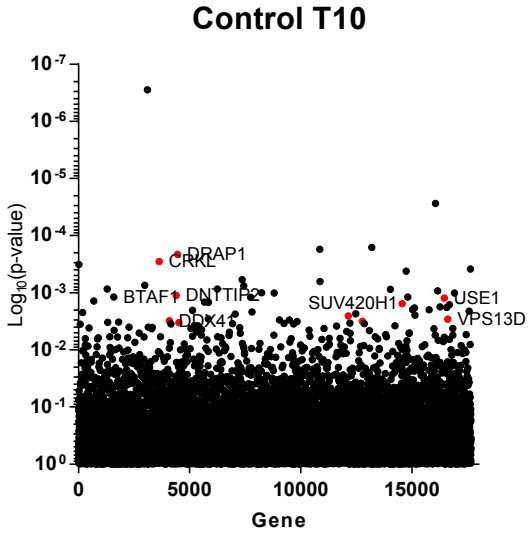


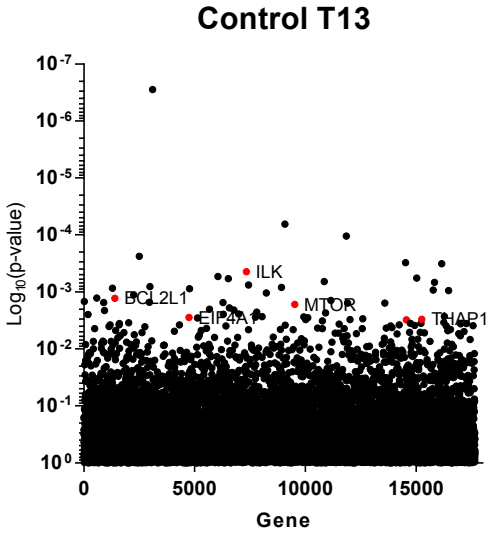
Figure 7. Average Log_2 Fold changes of all sgRNAs for known essential genes with representation of 3 or more at fold changes of less than -2 in HAP1 Control screen population.

To address these issues, several computational tools were developed to confidently distinguish between essential genes those that are not essential for cell survival. To enable the identification of essential genes for oxidative stress survival we employed MAGeCK, an analysis tool developed by Li et al.[60] MAGeCK uses an NB distribution to model the variance of individual sgRNAs and test their significance. From these individual tests, gene rankings can then be generated in accordance to their level of essentiality. MAGeCK was used to identify essential genes for all Control and peroxide-treaded time points (T_4, T_7, T_{10}, T_{13}, T_{16}) relative to the starting population (T_0). The results were filtered by fold change (less than -1) and p-value (less than 0.05) and the top 20 results selected. All control time points included previously identified essential genes with the total representation appearing to increase over the course of the screen. (**Figure 8**)


Control T4
Log10(p-value)
Gene
AASDHPPT
BRCA1
DDOST
ELP2
MTOR


Control T7
Log10(p-value)
Gene
AARS
CENPC1
CLRX3
POLR2L
SRP72
POLR2E


Control T10
Log10(p-value)
Gene
BTAF1
CRKL
DDX41
DRAP1
DNTTIP2
SUV420H1
USE1
VPS13D


Control T13
Log10(p-value)
Gene
BCL2L1
EIF2AK4
ILK
MTOR
TNAP1

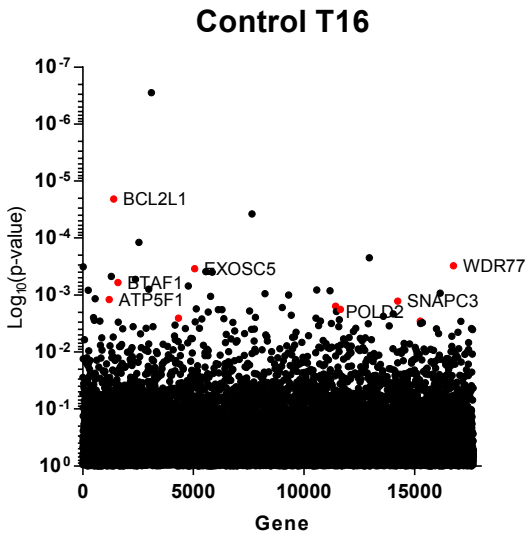
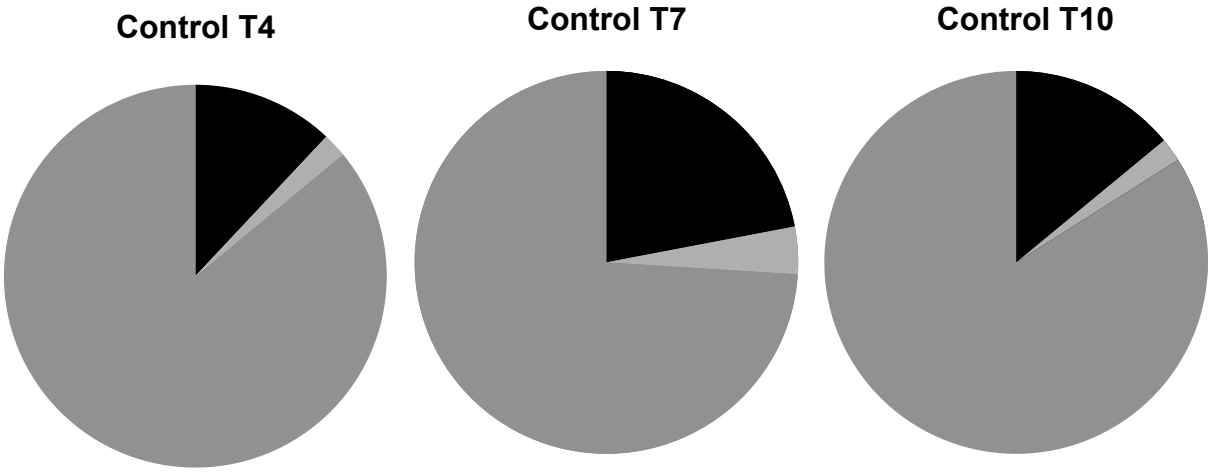


Figure 8. Previously identified essential genes identified as hits (shown in red) for each sampled time point of the HAP1 control screen.

To ensure that this was the case, the top 50 genes identified for each timepoint of the control screen were assessed to determine the proportion of genes deemed essential. From this, it became clear that the majority of these genes had not been previously identified as essential genes in prior published screens with the proportion of essential genes ranging from 15-25% from one timepoint to the next. (**Figure 9**)



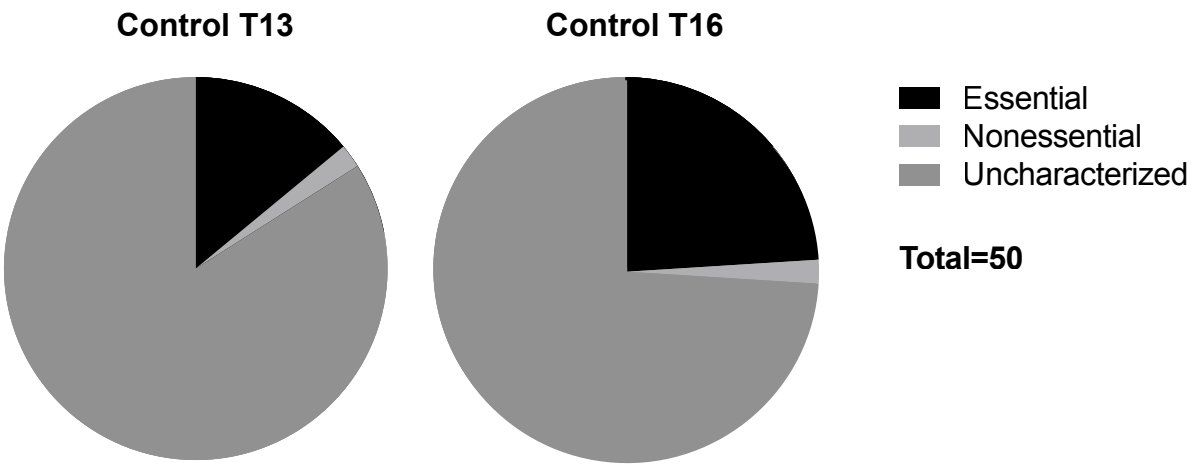


Figure 9. Proportion of essential to uncharacterized genes indentified at each timepoint of the HAP1 essential gene control screen.

Although the highest number of essential genes were found at the screen endpoint (T_{16}) there was no discernable trend suggesting that the number of essential genes identified increased with the duration of the screen. That the majority of genes identified by the screen have not been found to be essential suggests that the screen lacks sensitivity and that many of these genes may be nothing more than false positives. The list of essential genes identified by MAGeCK in the H_2O_2-treated population was compiled and purged of genes previously identified as essential or present in the untreated control population and the top 20 remaining hits reported as previously. (**Table 1**) Genes identified in the H_2O_2 population were aggregated and fed into GORilla[61] using the aggregated list of control gene hits as a reference background to determine essential biological processes involved in oxidative stress. From this, only a minimal enrichment for cellular aromatic compound metabolic processes (GO:0006725) was identified with no enrichment for either function or component. Although individual hits here identified may indeed be essential for survival under conditions of oxidative stress, these ontology results indicate that many of these identified hits may not be involved in the cellular response to oxidative stress and represent false positives.

Control					H_2O_2-Treated				
T4	**T7**	**T10**	**T13**	**T16**	**T4**	**T7**	**T10**	**T13**	**T16**
SOX6	EV12B	TSEN34	PPP5D1	**BLC2L1**	CRLF1	YPEL3	BCL3	C17orf67	ORAI2
MYRF	P2RX2	SAFB2	SPTBN4	KCNE1	C5orf52	C5orf52	FETUB	BLOC1S6	PRSS50
DDOST	MYRF	**DRAP1**	GGT5	CCDC41	HOXA10	NDUFB7	N6AMT2	SLC35A4	OR2A25
RERG	SLC31A2	STX16	GSX2	**WDR77**	TSEN34	WDR54	PEX3	C11orf71	GPR137
SREBF1	CEACAM6	KDELC2	P2RY2	**EXOSC5**	COL2A1	MAP4	RRAGB	LCE3A	COL2A1
TBC1D30	**CENPC1**	B2M	MCHR1	ELF2	CCL5	PCED1A	ZRSR1	HNRNPA3	SLC31A2
OR14K1	B3GNT2	TBC1D30	ELF2	PER2	PRKX	GALR3	TDRD15	OR5AP2	SAT2
SPATA2	LRFN1	**PSMG3**	CACHD1	MR1	SNX7	SEC22A	TRERF1	PRKAG1	PLA2G7
HIST1H1B	ORAI2	LCE3A	AMACR	**PKMYT1**	TRIM32	PIGP	GPM6A	BSDC1	NDUFAF1
ZNF474	RASL10A	CHTF18	PDZD7	MEIS3	BASP1	ZNF583	NLGN4X	IGLL1	BLC2
MIIP	**KANSL2**	TDRD15	USP10	FTH1	METTL7B	ARHGAP19	DHRS4L2	C12orf39	ARL9
LCE3A	**WDR77**	NDUFAF1	**BCL2L1**	ITPKA	IGF2BP1	PPOX	HIST1H4H	SLC25A33	PPP4R2
IFNE	LCE3A	ZNF836	**EIF4A1**	PLAA	MMACHC	RBM45	MORN1	KRTAP13-2	C6orf201
OR5D14	ADCK4	SDF2L1	NKX1-2	**ATP5F1**	OR2A25	ZBTB14	CCDC41	DNAJC19	COX7A1
CBY3	OR1L6	ODF3B	GULP1	**SNAPC3**	ZMAT4	RERG	PDE9A	KRTAP6-3	LILRA1
ACP1	**SF3B2**	**SNAPC3**	GPAT2	ALDH1L2	OR4C16	RAG2	ZXDA	TRIM32	VHLL
VENTX	GUCA2B	ACP1	TFE3	GPR45	CCDC169	PON3	HIGD1A	PTPLAD2	CNGA3
SUGT1	UROS	FTH1	PLEKHG4	HIPK1	TMEM246	PPP1CB	SLC25A33	ZDHHC11	DRD5
CCL15	OR5A2	ZNF223	STX16	LCE3A	C11orf73	DGCR6	VAPA	OR6C1	BCL2L
MAGOH	A1BG	C17orf51	ATP2B2	TRIM7	HNRNPA3	SMPD4	COA5	NIPSNAP3B	PRDX1
			CENPK						

Table 1. Top 20 essential genes for Control and H_2O_2-treated screen populations exhibiting Log_2 fold changes greater than -1 over all sampled time points (T_4, T_7, T_{10}, T_{13}, T_{16}).

1.5 Discussion

Herein, we attempted to systematically identify the essential genes of the cellular oxidative stress response using whole-genome CRISPR knockout libraries. Cas9-expressing human haploid HAP1 cultures were treated with a ramped concentration of H_2O_2 starting just below the experimentally determined[62] IC_{50} from 100-300μM and the population sampled at regular intervals alongside a control population to assess the proper functioning of the screen. The top 20 hits from each time point in the untreated control population revealed an increasing representation of core essential genes as previously identified by Hart et al.[33] indicative of proper screen function. The top prospective gene hits for the H_2O_2-treated population were determined after removing all high scoring candidates that were represented in either previously identified core essential genes or control screens. The aggregate list of these genes corresponded very little in their known functions or the processes in which they participate. This was unfortunate as our

original intention had been to broadly identify key components of the oxidative stress response. That we were unable to accomplish this may be due to the lack of sensitivity suggested by the prevalence of genes in the control screen that have not been shown to be essential. This might have resulted from a lack of statistical power, an issue that could be easily corrected by the inclusion of biological replicates for the screen. It may also be due to the inherent limitations of using hydrogen peroxide as genotoxic stressor. The size of the TKOv1 library necessitated that multiple large-scale cultures be maintained simultaneously. This placed limitations on how H_2O_2 could be introduced to the screen population. As it was technically infeasible to maintain a constant level of H_2O_2 over the course of the screen for lack of straightforward methods to constantly generate H_2O_2 or monitor it, simple bolus doses of peroxide at the indicated concentrations were introduced to fresh culture medium every 24hrs. This might have resulted in high localized concentrations of H_2O_2 that indiscriminately induced cell death irrespective of genetic vulnerabilities. The short duration of the stressor may also have contributed to the lack of representation of genes known to play a part in mediating oxidative stress survival. Given all of these limitations it is possible that only those genes whose loss induces the strongest sensitivity to oxidative insult could have been identified using this screen approach. Indeed, there are several top scoring genes that are known to play a role in cell survival. These include antiapoptotic factors BCL2 and 3. (**Figure 10**)

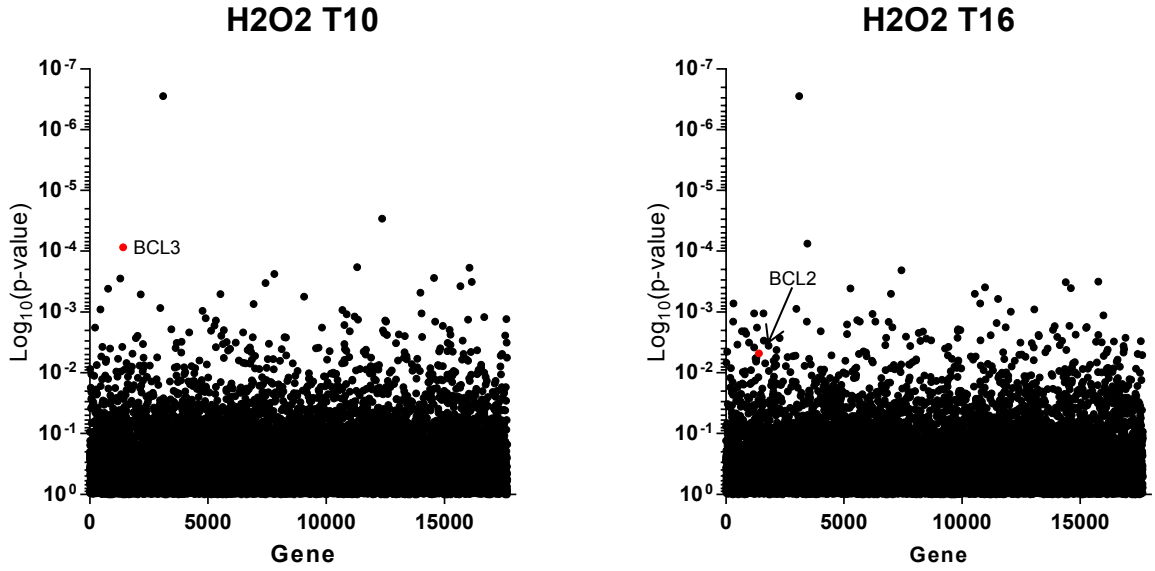


Figure 10. Identified essential genes candidates implicated in cell survival.

Other genes identified here as hits were directly implicated or indirectly associated with processes critical to cellular survival under stress conditions including cell growth (IGF2BP1, YPEL3, GPR137, SMPD4) (**Figure 11**), mitochondrial maintenance (SLC25A33, DNAJC19, NDUFAF1, COX7A1) (**Figure 12**), lipid biogenesis (PRKAG1, PTPLAD2) (**Figure 13**), DNA repair (HNRNPA3, RBM45, RAG2, PPP4R2) (**Figure 14**), and oxidative stress (BASP1, MMACHC, NDUFB7, PPOX, PON3, HIGD1A, VHLL). (**Figure 15**)

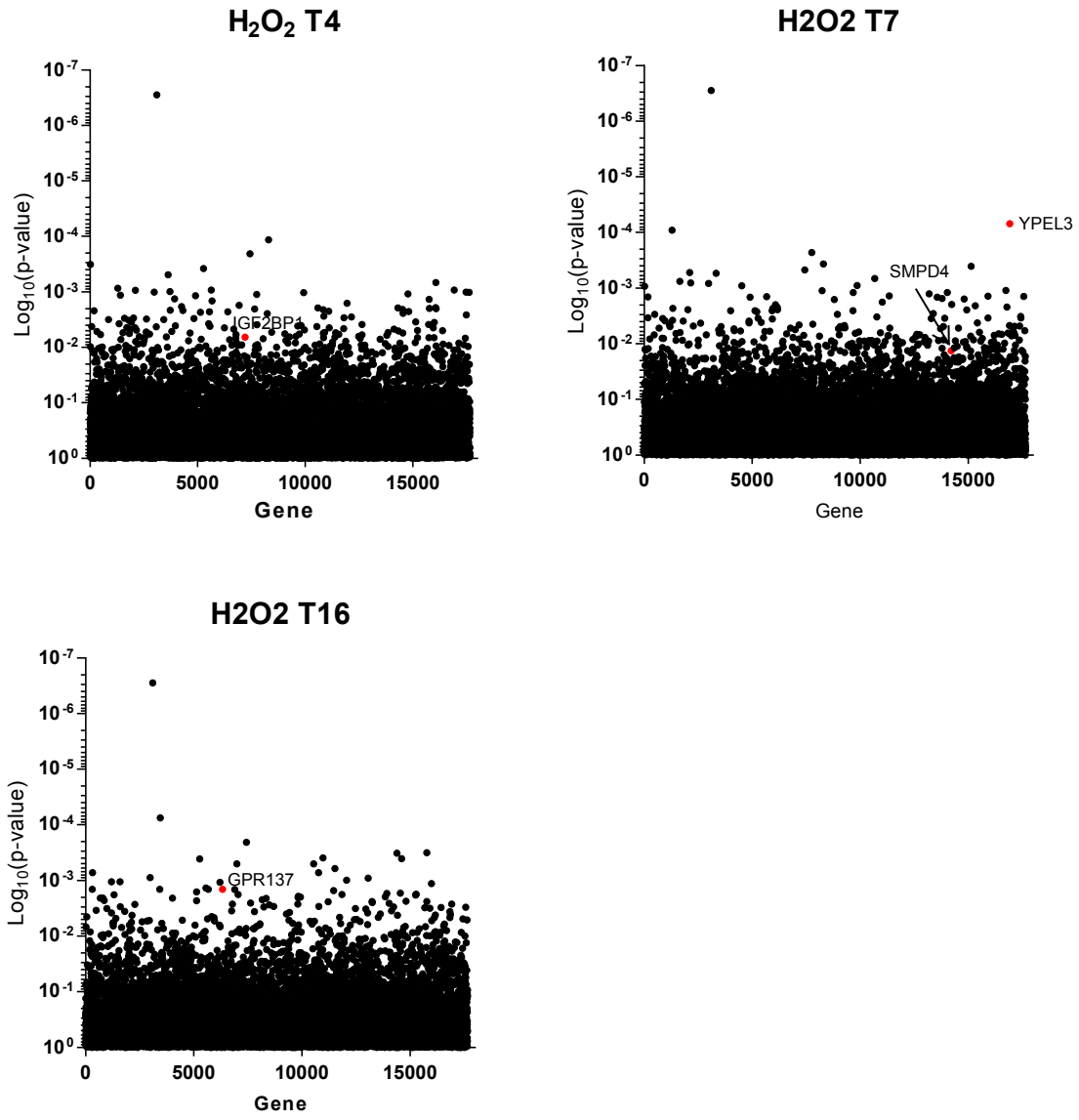


Figure 11. Identified essential gene candidates implicated in cell growth and proliferation.

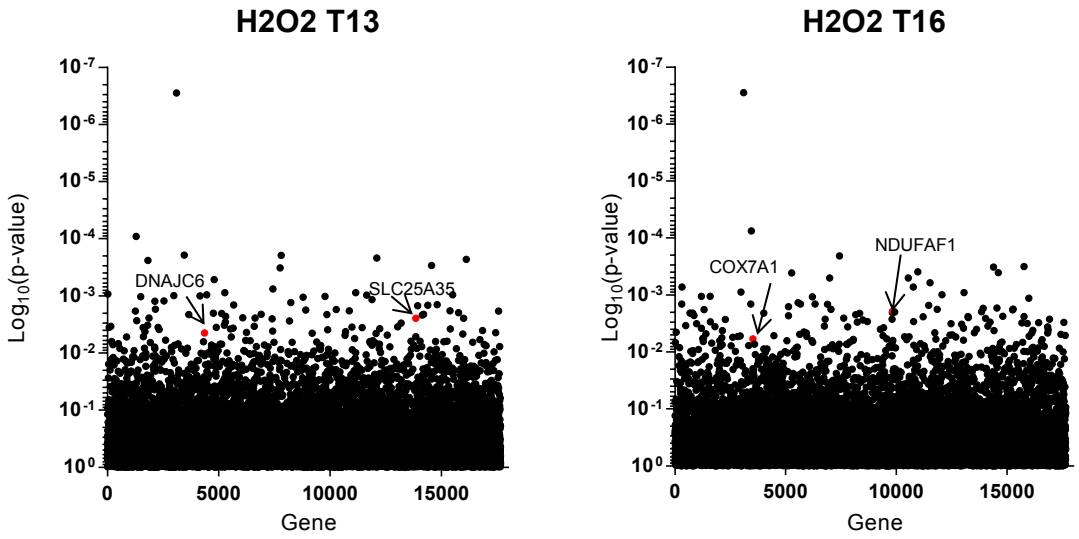


Figure 12. Identified essential gene candidates implicated in mitochondrial maintenance.

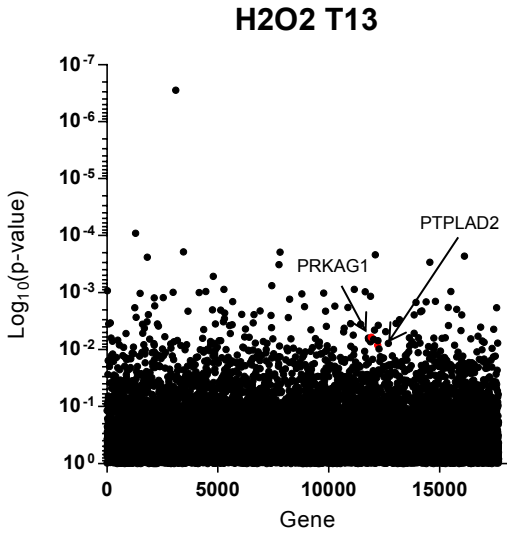


Figure 13. Identified essential gene candidates implicated in lipid biogenesis.

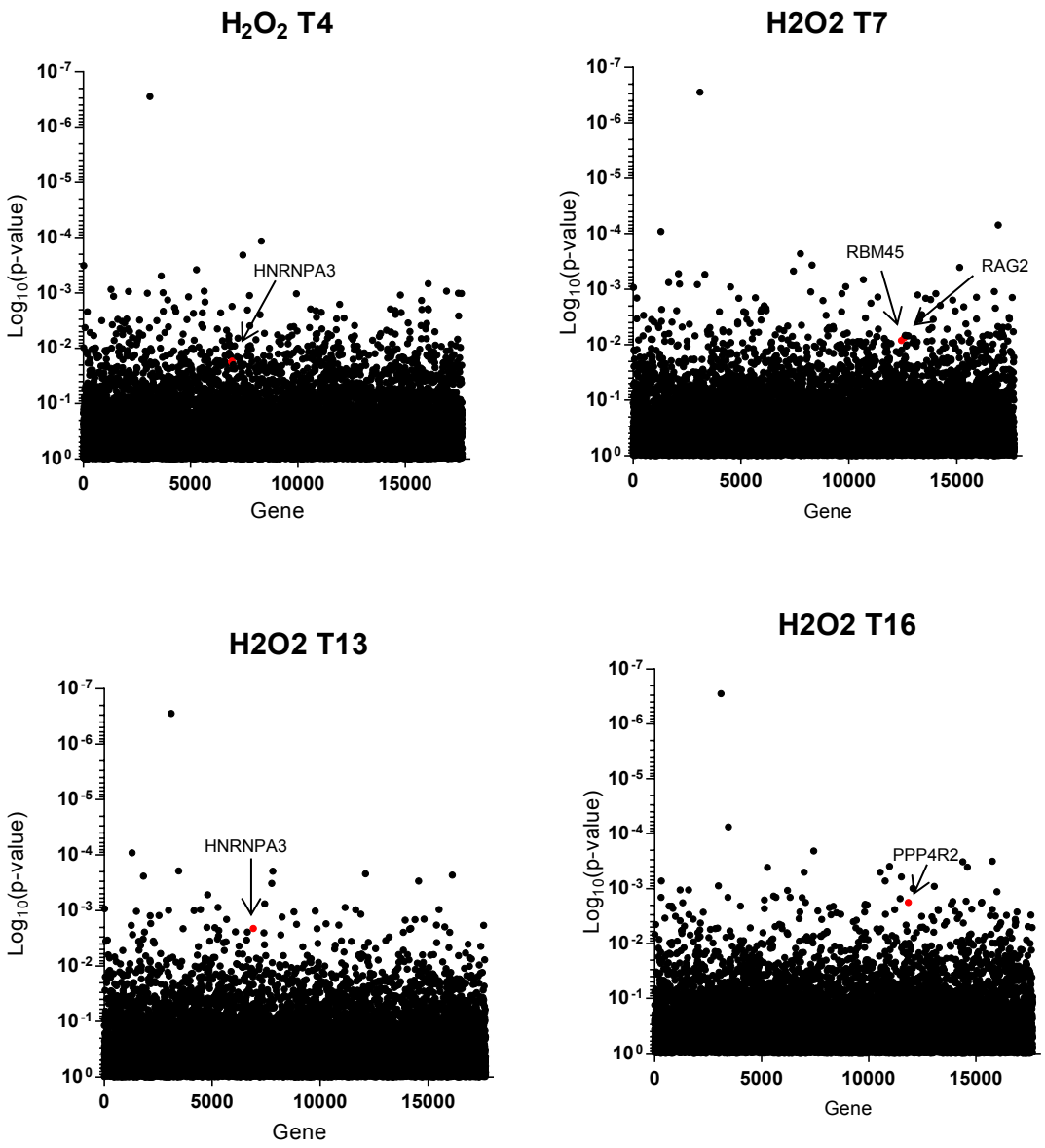


Figure 14. Identified essential gene candidates implicated in DNA damage response and repair.

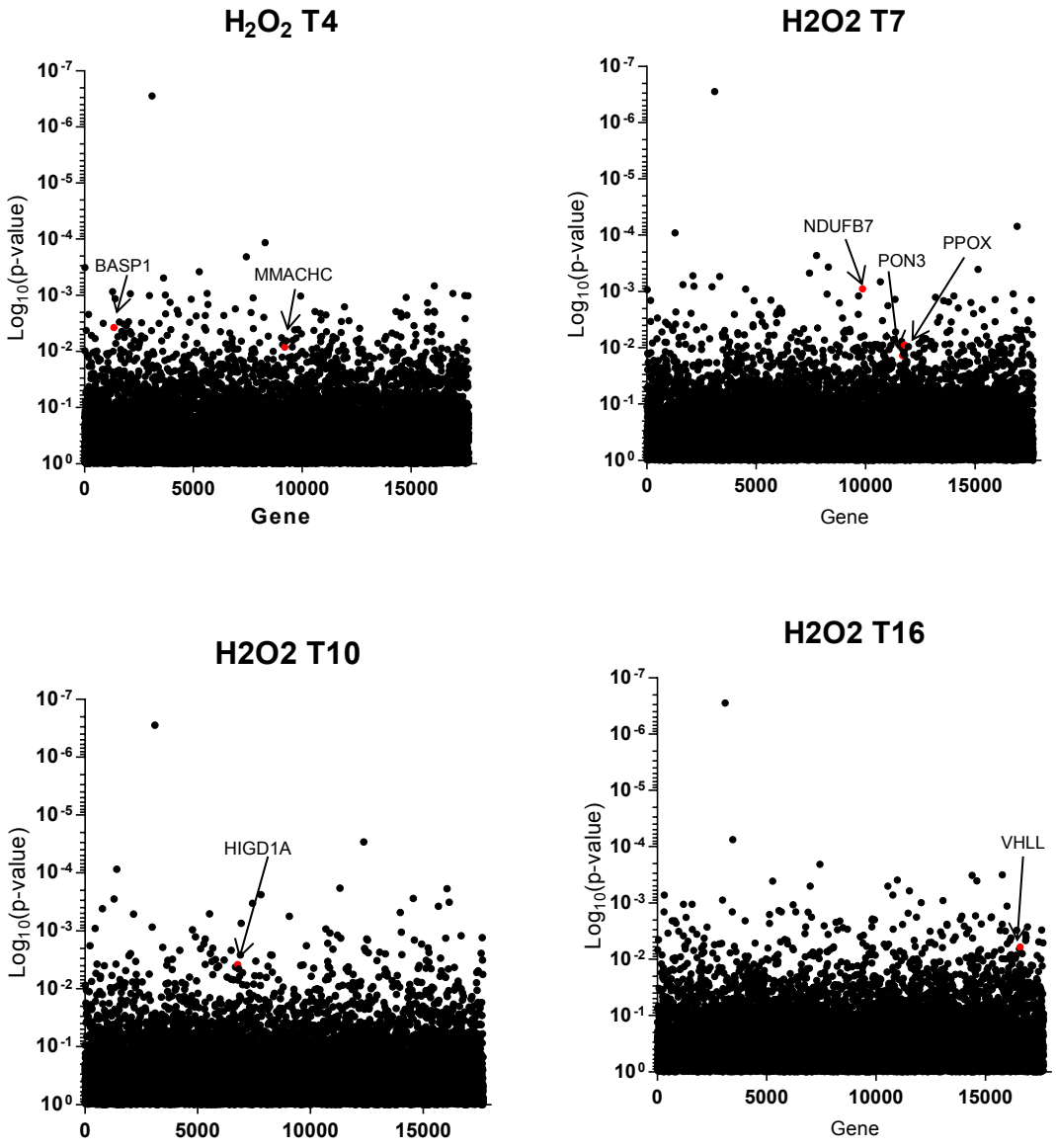


Figure 15. Identified essential gene candidates implicated in the cellular oxidative stress response.

Of particular interest here are MMACHC and VHLL, two genes with distinctly different functions that have been implicated in oxidative stress but whose function is completely unknown. Metabolism of Cobalamin C (MMACHC) has been proposed by homology to play a role in the intracellular processing and trafficking of cobalamin. Vitamin B_{12} deficiency has been associated with increased oxidative stress and has been proposed to mitigate oxidative stress by scavenging ROS or maintaining homocysteine metabolism.[63] Whether or not vitamin B_{12} and, by association, MMACHC function in either of these capacities is, as of now, unknown. For DNA repair, RBM45 has been implicated as an important regulator of the DNA damage response[64] whereas HNRNPA3 has been implicated in telomere stability among other functions.[65,66,67] Both of these proteins though have been implicated in onset of ALS and so might represent interesting candidates for further study. Finally, DNAJC19 is a chaperone protein associated with mitochondrial maintenance and repair that is frequently mutated in a number of diseases linked to oxidative stress including cardiac dysfunction and Parkinson's disease.[68] There are also a number of genes represented as top hits that, as of yet, have no known function. These include uncharacterized GPCRs (OR2A25, OR4C16, OR5AP2, OR6C1), zinc-finger proteins (ZNF583, ZFP69B), and ORFs (C5orf52, C17orf67, C11orf71, C6orf201). (**Figure 16**) It would be of interest to see if these in fact do play a role in the survival of the cell under conditions of oxidative stress. In all these cases, individual validation of the genes would need be conducted to confirm the screen results before further work could be proposed to characterize their functions.

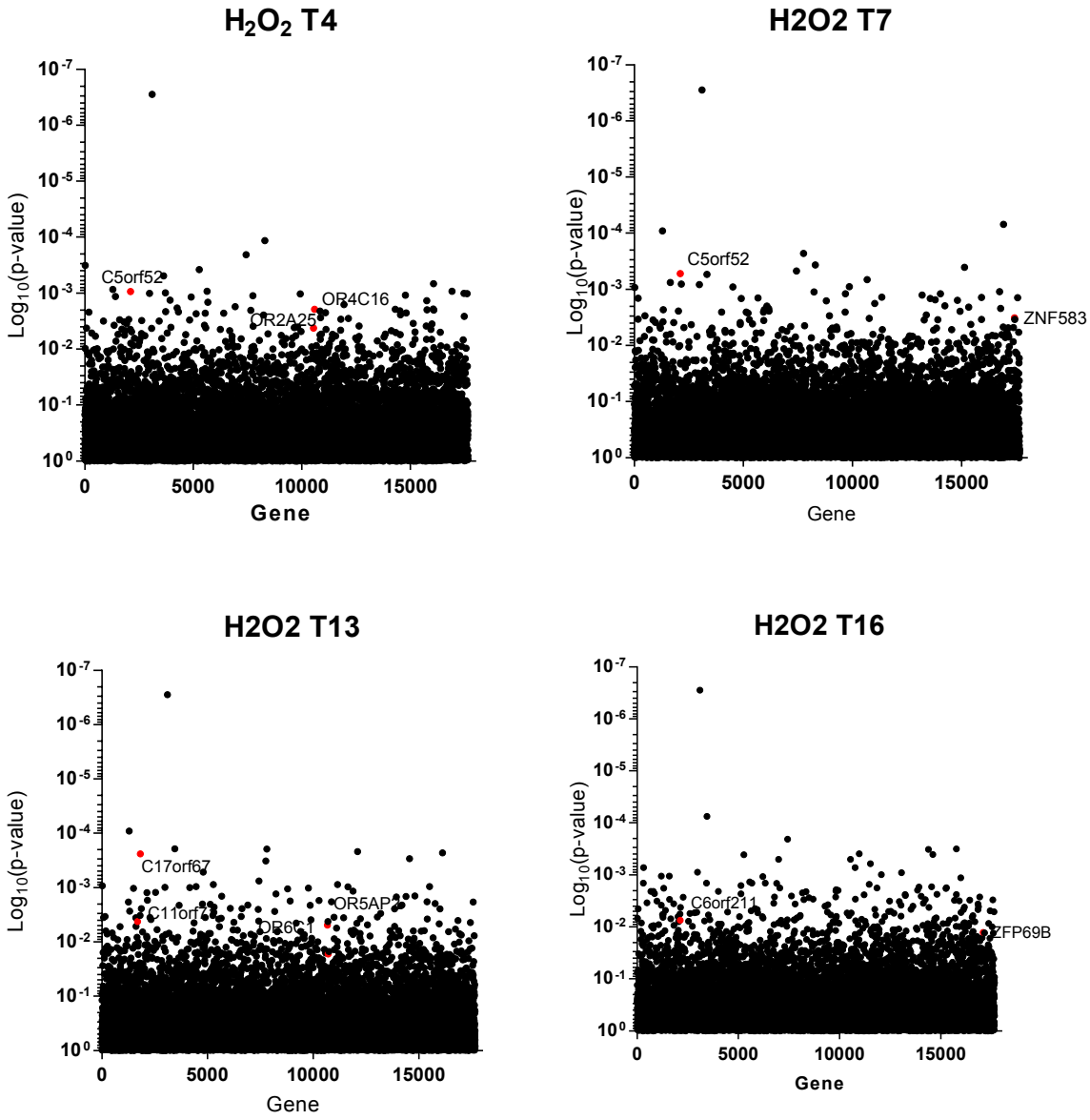


Figure 16. Identified essential gene candidates with uncharacterized functions.

A recent and more comprehensive study published by Dubreuil et al.[69] however, tempers our expectations on the validity of these results. Here, CRISPR KO screens were used in combination with shRNA screens to identify mediators of oxidative stress. Identified hits were then validated with a second round of screens using CRISPR KO and CRISPRi sublibraries

followed by single hit growth assays. Similar to what was observed in our own screens, the identified genes included both known and unexpected mediators of oxidative stress. From their screens, they found that PRDX1, a peroxiredoxin, sensitized cells to oxidative stress whereas KEAP1, a negative regulator of the master transcription factor for oxidative stress, was protective. This was also observed in our screens at the final timepoint under the highest levels of oxidative stress (T_{16}). However, our screens failed to identify any of the other major antioxidative genes identified including the antioxidative protein Catalase (CAT), the iron homeostasis regulator IREB2, members of the peroxisomal import pathway (PEX), and the pentose phosphate pathway mediator PGD. That some of the genes identified in these screens were also found in our own but so many others missed supports our supposition that the cells in culture were not sufficiently stress with H_2O_2 to promote genetic selection. In their screens, Dubreuil et al.[69] used multiple treatments at the experimentally determined IC_{50} of their cell models, allowing the cultures to recover for several days before starting another round of treatment. This suggests that our single treatment was not of sufficient duration to induce genotoxic stress and that many of the genes identified are likely spurious. This issue would necessitate that another screen be performed using multiple treatments of a fixed rather than a ramped H_2O_2 concentration. Dubreuil et al.[69] along with many others [38,70,71] have demonstrated the usefulness of performing identical screens with different methods in order to more reliably identify essential genes for the process in question. Given this, it may be prudent to combine whole genome CRISPR KO screens with CRISPR activation screens (CRISPRa) to identify previously unknown mediators of the oxidative stress response. This would provide further experimental support for the published results of Dubreuil et al.[69] and may, with the addition of CRISPRa, lead to the identification of further genes that have not yet been uncovered.

1.6 Conclusions

Reactive oxygen species play critical roles in cellular function both normal and abnormal. As such, numerous proteins that deal with their regulation are encoded in the genome, many with roles that have yet to be defined. We have attempted to employ the recently developed CRISPR-Cas9 knockout screen libraries to identify these unknown regulators. While a preliminary screen on a control population successfully identified some previously uncovered essential genes a lack of sensitivity was observed and performing the screen on a population treated with H_2O_2 likewise yielded sparse and largely unexpected putative gene hits. A recently published study employing nearly identical methods to our own supported very few of these and identified many more that were entirely absent from our screens. This undermines whatever confidence we might have in the genes here identified playing any part in regulating cellular ROS. It also suggests that this failure could be corrected with a simple modification to how selective pressure was applied to the screen cultures along with the inclusion of screen replicates. Making these corrections and employing a CRISPRa whole genome screen to complement the CRISPR KO screen would amend the faults of our previous effort, lend support to the comprehensive work of Dubreuil et al.[69], and provide an opportunity to uncover further genes involved in the cell's response to oxidative stress.

Chapter 3

Uncovering Functional Redundancy in DHHC-family Palmitoyltransferases using CRISPR-Cpf1 Duplex Genomic Screens

3.1 Introduction

3.1.1 Protein Lipidation in Cell Biology

The post-translational modification of newly synthesized proteins is a characteristic feature of eukaryotic cell biology. Many examples of these modifications have been catalogued ranging from the addition of small molecules that occurs in phosphorylation, acetylation, methylation, and N-glycosylation to whole proteins as in ubiquitinylation and SUMOylation.[72] Many of these modifications are functionally reversible, with the addition and removal of each modification conducted by distinct groups of 'writer' and 'eraser' proteins respectively.[73]The presence or absence of these modifications play critical roles in protein function by permitting tight control over subcellular localization, protein-protein interactions, and enzymatic activity.[74] Lipidation, the post-translational modification of proteins with acyl fatty acids, accomplishes this by promoting association with the cellular and subcellular membranes and includes both reversible and irreversible modifications.[72] Of all of the observed lipid modifications so far only palmitoylation, the S-acylation of cysteine residues with palmitate, has been shown to be reversible.[73] The addition of palmitate is carried out by a large family of membrane-associated

DHHC protein acyl transferases (PATs) and removal by less well understood protein thioesterases such as acylprotein thioesterase 1 (APT1) and 2 (APT2) and the ABDH17 proteins.[74] Cycling between palmitoylated and depalmitoylated states catalyzed by PATs and thioesterases respectively permit close control of protein subcellular localization and, as a consequence, protein function.[75] (**Figure 17**) The importance of these palmitoylation cycles to normal cellular function is highlighted by the consequences of their dysfunction which has been associated with a wide assortment of pathologies including schizophrenia, developmental defects, and cancer.[76]

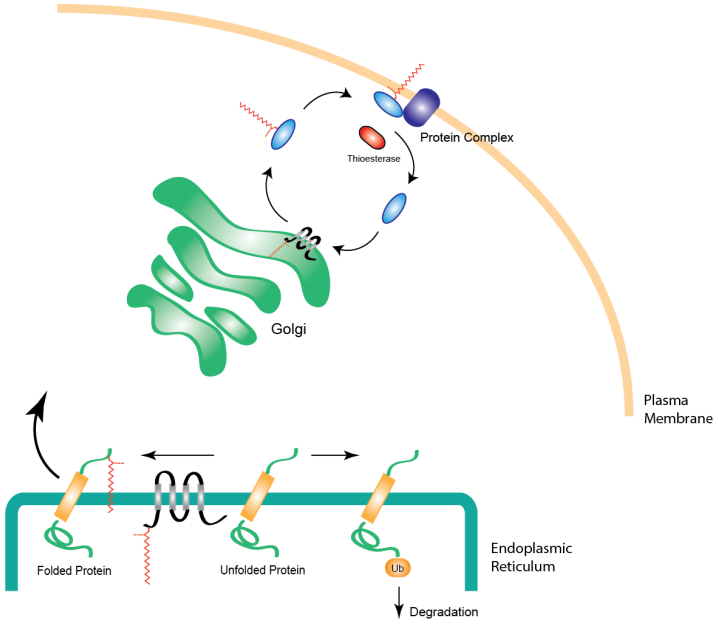


Figure 17. The varied roles of DHHC-family palmitoyltransferases. Palmitoylation of protein substrates by DHHC PATs promotes association of the target with the membrane and promotes movement through the secretory pathway. Palmitoylation my also promote stability of the target protein or specific protein-protein interactions.

3.1.2 The DHHC Family Acyltransferases

The reversible addition of palmitate to protein cysteine residues is carried out by a large family of DHHC PATs, so named for the conserved Aspartate-Histidine-Histidine-Cysteine catalytic

motif found in all of its members.[77] First identified in yeast, DHHC PATs have been found across multiple plant and animal species with a total of 23 distinct genes found in mice and humans respectively.[78]All DHHC PATs are membrane associated and share a conserved core of 4-6 transmembrane domains with a cysteine-rich domain containing the catalytic DHHC motif on an cytosolic loop.[79] Beyond this core structure however, DHHC PATs exhibit only limited sequence homology.[80] DHHCs catalyzed the transfer of palmitate to their associated substrates by two-step ping-pong mechanism.[81] The first step involves the formation of an acyl-intermediate between the palmitoyltransferase and palmitoyl-CoA through autopalmitoylation followed by the transfer of palmitoyl group to the substrate in the second step.[82]The cysteine of the DHHC catalytic motif is essential for both autopalmitoylation and transfer of the palmitoyl group whereas loss of the first histidine only prevents the latter.[83]

3.1.3 Inter- and Intracellular Distribution of DHHC-Family Acyltransferases

DHHCs PATs have been found in association with the membranes of many subcellular compartments including the Endoplasmic Reticulum (ER), Golgi apparatus (GA), plasma membrane (PM), and transport vesicles. The work of Ohno et al.[84] was the first comprehensive effort to determine the subcellular distribution of DHHC PATs. From this it was discovered that DHHCs 1,6,11,13,14 and 19 were found to stably associate with the ER, DHHCs 3,4,7,8,15,17, and 18 with the GA, and DHHCs 5, 20, and 21 with the PM. DHHCs 2, 9, 12, and 22 by contrast were found in both the ER and GA. Subsequent studies have revealed alternate localization patterns for some of these DHHCs and may be reflective of differences in cell type.[85] DHHC 1 was found to localize to Rab5+ early endosomes[86] in neurons, and both the GA and ER in HeLa cells.[87] Localization of DHHC2 to both the ER and GA was reported as previously[88,89] along

with evidence of localization to the PM[90,91,92] and endosomes.[92] DHHC4 was found to localized to the ER[93] in addition to the GA.[94,95] DHHC 11, like the closely related DHHC1, was also found to localize to endosomes.[96] DHHC13 was found to localized predominately to the GA[96,97,98] but also endosomal vesicles.[97,98]DHHC14, like the related DHHC9, has been found to localize to the GA.[99]The localization of DHHC16, previously undetermined, was established at the ER.[88,100] Subsequent work on DHHC19 revealed localization at the GA.[101] Similar to DHHC2, DHHC20 showed dispersed localization in multiple subcellular compartments.[102] DHHC21 was found to alternately localize to the GA.[103] Absent from previous work, DHHC23 was subsequently shown to localize to the GA and trans-Golgi network.[104] With some exceptions, the majority of DHHCs stably partition between the ER, GA, and PM, the major subcellular organelles involved in the classic secretory pathway. Movement of substrate proteins through the secretory pathway is determined by the competing actions of PATs and thioesterases, with palmitoylation by the resident PATs in each subcellular organelle promoting membrane association and driving them forward through the secretory pathway. Given this, it may be presumed that every substrate has a specific PATs or set of PATs within each organelle with which it selectively associates for palmitoylation. With a larger number of PATs residing in the GA than the ER and PM respectively, this would indicate that PATs localized to the former are responsible palmitoylating a narrower range of substrates than the latter; that GA resident PATs exhibit inherently greater substrate selectivity whereas those of ER and PM are more promiscuous. The specific features of each PAT that determine its subcellular localization are still largely unknown. There is high degree of conservation within the structural core, containing the transmembrane and the cysteine-rich catalytic domains of all DHHCs. This would indicate that the structural determinates of subcellular localization and, possibly, substrate preference reside within the N-

and C-terminal regions where nearly all variability between DHHCs is found. In support of this, Greaves et al. found that the C-terminal domain of DHHC2 was found to be necessary for its localization to the PM.[92] ER-resident DHHCs 4 and 6 were found to contain C-terminal lysine-based sorting signals that were essential for their retention and, indeed, were sufficient to alter the localization of non-ER resident PATs to this organelle.[93] Aside from these examples, no subsequent work has been done to investigate the variable N- and C-terminal domains of other DHHCs and how they influence their subcellular localization. Surprisingly, a systematic investigation of these questions has not been undertaken and should be done to resolve them.

3.1.4 Substrate Preference and Functional Redundancy within DHHC-Family Acyltransferases

Of the questions surrounding the function of palmitoyl transferases, none is more intriguing than that of substrate preference. At the heart of this question is an apparent paradox. All of the model organisms surveyed have been found to contain multiple, distinct PATs encoded within their genome, with a greatest number being found in higher organisms, humans and mice with 23 respectively. This would seem to imply a high degree of substrate specificity, each PAT having a defined set of proteins that it palmitoylates to exclusion of others. This would be in keeping with the large number of DHHCs found in many organisms; if DHHCs PATs did not have some degree of inherent substrate specificity there would no selective advantage to retaining them and other evolutionary time their numbers would be whittled down to an essential few, maybe a single PAT for each organelle of the secretory pathway. This rationale was driving force behind initial efforts to identify the unique substrates of each PAT and the principles that dictate their specificity. An understanding of these could, in principle, guide the development of

pharmacological compounds targeted to specific DHHCs involved in particular disease-causing pathways, while leaving the rest to function unperturbed. Indeed 2-bromopalmitate, the only PAT-specific inhibitor then developed, targeted the catalytic domain and therefore operated indiscriminately across all PATs to impede palmitoylation.[105] The work of Ohno et al.[106] and subsequent efforts however yielded results contrary to this expectation. Screening of DHHCs PATs against substrates of interest by H^3-palmitate labeling reveled a propensity of certain DHHCs to palmitoylate these substrates over others. Knockdown or knockout of these same PATs however, did not abolish palmitoylation of these same substrates *in vitro* or greatly impede their regular function. This was likewise demonstrated by Roth et al. where single knockouts of yeast DHHCs failed to alter the palmitoylation state of multiple palmitoylated proteins, an outcome that only occurred when five or more DHHCs were simultaneously inactivated.[107] This would imply an unexpected degree of functional redundancy between PATs. The extent of this functional redundancy and the reasons for it remain unresolved.[108] DHHCs are often grouped into subfamilies on the basis of sequence similarity. (**Figure 18**) Presumably, DHHCs within these subfamilies possess similar structures, particularly at the N- and C-terminal domains believed to be responsible for intracellular localization and/or substrate recognition. This would result in the palmitoylation of a similar panel of substrates either by either driving localization to specific subcellular compartments or association with specific features of the substrate. The members of certain DHHC subfamilies are known to have relatively complex structural features outside of the cysteine-rich catalytic domain. DHHCs 13 and 17 are known to have ankyrin-repeat binding domains while DHHCs 5, and 8 possess PDZ-binding motifs and DHHC6 an SH3-binding motif.[109] These have all been shown to promote association with and palmitoylation of particular substrates. Other DHHCs such as Golgi resident DHHC3 and 7 however seem to

lack such structural features and do not seem to exhibit much discrimination between the substrates that they palmitoylate. This would suggest that DHHC PATs may be subdivided into two groups 1) those that exhibit lower activity to all possible substrates but a high degree of selectivity toward a few and, 2) those that have high activity toward all substrates but exhibit little to no selectivity between them. Localization may also play an important role in substrate specificity as DHHCs can only palmitoylate substrates that occupy the same subcellular space. DHHCs within the same cellular organelles may then be expected to palmitoylate the same substrates by simple proximity. This would result in a high degree of functional redundancy between GA-resident PATs and a lesser degree in those of the ER and PM. It is probable that both of these considerations play a role in determining the substrate panel of any given DHHC such that all DHHCs within a given organelle may palmitoylate any substrate within their immediate vicinity but will associate more with those that possess complementary structural motifs. Therefore, the loss of any given DHHC may be compensated to some degree by the others that occupy the same space within the cell and mask any resulting phenotype. The expression of DHHCs varies from one tissue to another[92] and so the effects of losing certain DHHCs may be more pronounced than others, particularly if they possess substrate recognition domains. It can be predicted then that some degree of functional redundancy may exist between PATs localized to the same organelle and even more so if they derive from the same structural subclass.

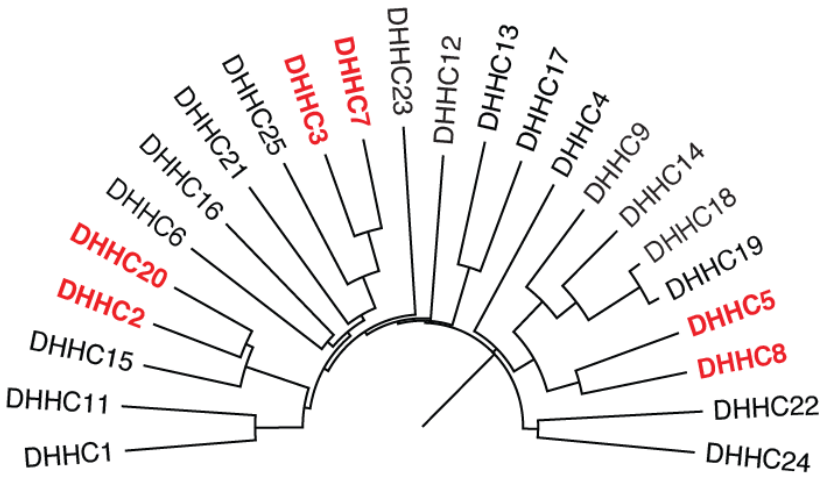


Figure 18. Sequence homology relationships between DHHC-family palmitoyltransferases. The high degree of relatedness between DHHCs 2/20, 3/7, 5/8, and 18/19 denote distinct subfamilies with presumed similarity in function and substrate preference.

3.1.5 Ras GTPases in Normal and Aberrant Cellular Processes

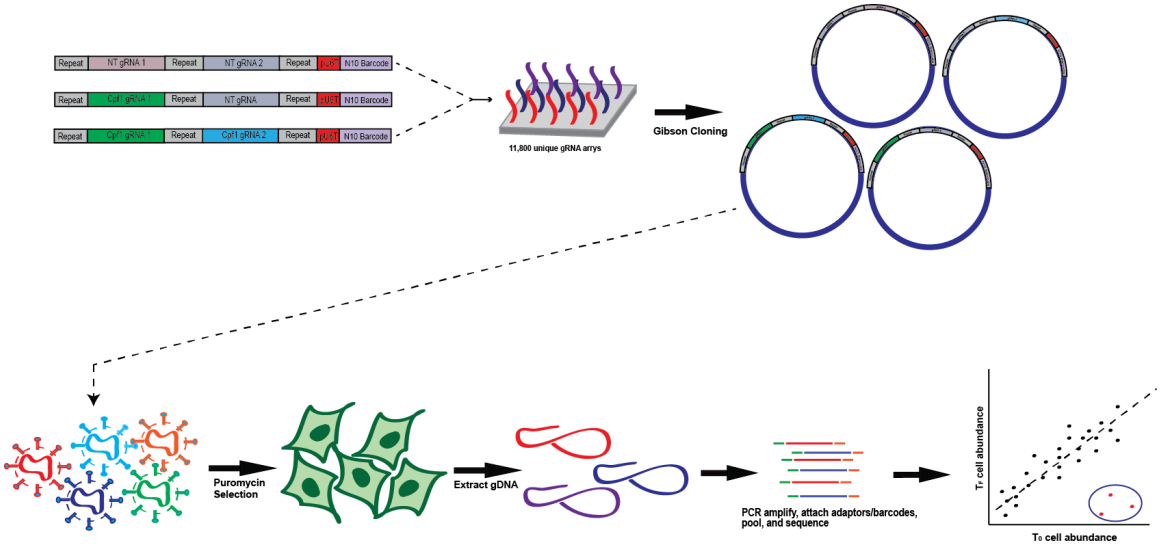
The Ras proteins make up a superfamily of over 150 small, monomeric GTPases that are well conserved in eukaryotes.[110] Members of the Ras-family GTPases function as nodes in signaling pathways, interacting with multiple downstream effectors upon activation by external stimuli, that, in turn, regulate signaling networks that control the expression of genes involved in cell proliferation, differentiation and survival.[111] The Ras GTPases are therefore essential drivers of normal cellular proliferation. Given their critical role in this process however, it is unsurprising they are all to often the culprits in the rampant cell growth that is characteristic of aggressive malignancies.[112] Ras activation is triggered by the exchange of bound GDP for GTP, promoting a conformational change that increases the affinity of Ras for its downstream effectors.[113] This state of activation persists until the bound GTP is hydrolyzed to GDP. Ras, by itself, is intrinsically poor at carrying out either of these actions and so relies on associated Guanine-nucleotide-exchange factors (GEFs) and GTPase-activating proteins (GAPs) to carry promote activation and inactivation respectively.[114] Aberrant function of Ras or any of its associated

GEFs and GAPs can result in inappropriate activation of Ras and its downstream effectors. Specific mutations in Ras that promoted a constitutively active state by preventing GTP exchange were responsible for the initial discovery of Ras and recognition of its role in cancer progression.[115] All of the primary Ras isoforms require membrane localization in order to access their downstream targets and drive proliferative signaling.[116] This assertion is supported by the extensive lipid post-translational modifications that have been found in association with Ras. Canonical Ras isoforms all contain C-terminal cysteine motifs that serve as the recognition signal and attachment point for C15 or C20 lipids by farnesyl or geranylgeranyl transferases respectively.[117] This modification permits some degree of membrane interaction but has been found to be insufficient for the stable membrane association required for trafficking to the plasma membrane.[118] This association with the membrane and ultimately the function of Ras is controlled by antagonistic cycles of palmitoylation/depalmtoylation[119] and this dependence suggests that DHHC acyltransferases may play an essential role in Ras-dependent proliferation. Ras isoforms NRas and HRas have already been shown to be substrates for DHHC9, undergoing palmitoylation at the Golgi before trafficking to the plasma membrane.[120] Furthermore, Ras isoforms are unable to drive cellular proliferation when made palmitoylation-deficient either by genetic mutation or by chemical inhibition of DHHC palmitoyltransferases.[121,122] This indicates that there are likely many other Ras isoforms and associated proteins that are substrates for acyltransferases.[123]

3.2 Research Justification and Objectives

The question of functional redundancy between DHHC PATs remains unresolved. To address this, we aim to employ a pooled genetic screen approach targeting all known DHHC genes. It has

been previously shown that single knockout or knockdown of individual DHHCs results in a null phenotype due to compensation by the remaining unaffected DHHCs. In order to get around this issue we use the CRISPR-Cpf1 system paired with randomized duplex crRNA arrays targeted to all known DHHCs. (**Scheme 2)** Should any synergy or functional redundancy exist between a pair of DHHCs duplex arrays containing only a single targeting sequence for either PAT should yield a much less noticeable effect than an array containing crRNAs targeted to both. Libraries were constructed for both human and mouse genetic models following the methods developed by Chow et al.[124] Screens were then performed using these libraries in a general human haploid cell model and a more strictly controlled Ras-dependent mouse cell model. The human HAP1 cells are a nearly haploid cell line and so should prove more sensitive to genetic knockout. They will be used to validate the screen methodology prior to conducting screens in Ras-dependent cell models. The mouse MEF cell lines, meanwhile, have been engineered so as to be dependent of a single Ras isoform for their survival. As such, any disruption in the signaling pathway of that specific Ras isoform will cause an arrest in the cell's growth. Since Ras and many of its affiliates require palmitoylation to properly function, the loss of any PATs responsible for carrying this out will be lethal to any cell that has been made deficient in them. These Rasless MEFs also provide us with a unique opportunity to identify other PATs involved in Ras-dependent growth beyond Ras itself, something that has not yet been attempted. These screens should provide a good basis with which to investigate these fundamental questions surrounding the function of palmitoyltransferases.



Scheme 2. Methods to be used for Cpf1 library synthesis and subsequent pooled KO screen.

3.3 Methodology

3.3.1 Cell Culture

Isogenic Ras-dependent MEF cell lines (HRas, NRas, KRas4a and KRas4b) were obtained from the National Cancer Institute (NCI) and cultured in Dulbecco's Modified Eagle's Medium (Gibco) supplemented with 10% v/v Fetal Bovine Serum (HyClone) and 1% v/v penicillin/streptomycin (Gibco). All cell lines were maintained under selection (Puromycin 2.5µg/mL or Blastocidin, 4µg/mL) to ensure continued Ras-dependency. HAP1 cells were obtained from the laboratory of G. Van der Goot (Ecole Polytechnique Fédérale de Lausanne) and cultured in Iscove's Modified Dulbecco's Medium (Gibco) supplemented with 10% v/v Fetal Bovine Serum (HyClone) and 1% v/v penicillin/streptomycin (Gibco). HEK293T cell lines were maintained Dulbecco's Modified Eagle's Medium (Gibco) supplemented with 10% v/v Fetal Bovine Serum (HyClone) and 1% v/v penicillin/streptomycin (Gibco).

3.3.2 Design and Synthesis of mDHHC Arrays

AsCpf1/LbCpf1 sgRNAs using the TTTN protospacer adjacent motif (PAM) were identified within the coding sequence of all major zDHHC genes contained within the mouse genome using the CRISPR design tools provided by Benchling. All possible sgRNAs for each gene were compiled and four sgRNAs selected per gene with preference given to those located 5' to the conserved cysteine-rich catalytic domain and/or exhibiting a high off-target score (i.e those that exhibit low probability of cutting elsewhere in the genome) for a total of 96 DHHC-targeted sgRNAs. To this list of targeting sgRNAs was added 52 non-targeting control sgRNAs derived from those used by Chow et al.[124] Common 5' and 3' homology arms along with spacers containing the Cpf1 direct repeat were appended to all sgRNAs within the list and randomized to generate all possible combinations of targeting and control sgRNAs. Duplicate sequences (sgRNA 1 + sgRNA 1) and mirror duplicates (where sgRNA 1 + sgRNA2 = sgRNA2 + sgRNA 1) were removed to give a list of 10,878 unique duplex arrays. To these DHHC-specific arrarys was added 91 arrays targeting the primary Ras isoforms (NRas, HRas, KRas4a, and KRas4b) bringing the final library to 10,969 unique duplex arrarys. These were submitted to CustomArray Inc. for synthesis.

3.3.3 Design and Synthesis of hDHHC Arrays

AsCpf1/LbCpf1 sgRNAs using the TTTN protospacer adjacent motif (PAM) were identified within the coding sequence of all major zDHHC genes contained within the human genome using the CRISPR design tools provided by Benchling. All possible sgRNAs for each gene were compiled and four sgRNAs selected per gene with preference given to those located 5' to the conserved cysteine-rich catalytic domain and/or exhibiting a high off-target score (i.e those that exhibit low probability of cutting elsewhere in the genome) for a total of 92 DHHC-targeted

sgRNAs. To this list of DHHC-targeted sgRNAs was added 52 randomly generated non-targeting control sgRNAs. Random 20-bp oligonucleotide sequences were generated using RSAT and the resulting output purged of any redundant sequences using the Sort and Unique lines tools from Galaxy. The output was subsequently uploaded to Cas-OFFinder[125] and all sequences exhibiting fewer than 4bp mismatches to regions of Homo sapiens hg38 reference genome removed. The 52 non-targeting control sequences were then selected randomly from this filtered list. Additionally, crRNAs against genes selected from the review of essential genes produced by Telenti et al.[126] (hEIF3B, hDICER1, hPOLR2A, hNRF1; 3 sgRNAs per gene) were also included as controls to evaluate screen function. Common 5' and 3' homology arms along with spacers containing the Cpf1 direct repeat were appended to all sgRNAs within the list and randomized to generate all possible combinations of targeting and control sgRNAs. Duplicate sequences (sgRNA 1 + sgRNA 1) and mirror duplicates (where sgRNA 1 + sgRNA2 = sgRNA2 + sgRNA 1) were removed to give a list of 12090 unique duplex arrays. These were submitted to CustomArray Inc. for synthesis.

3.3.4 Cloning of mDHHC/hDHHC sgRNA Libraries

The parent lentiviral vectors pRC49 (puromycin resistance, firefly luciferase, nuclear EGFP) and pRC11 (puromycin resistance) for cloning of the mDHHC and hDHHC libraries respectively were obtained as a gift from Sidi Chen (Addgene # 123363) and digested with Esp3I according to the following conditions:

-5µg Plasmid DNA
-3µL FastDigest Esp3I (ThermoFisher #FD0454)
-4µL FastAP (ThermoFisher #EF0651)
-5µL 10x FastDigest Buffer
-ddH_2O to 50µL

Thermocycler conditions: 37°C for 60 minutes.

The plasmid digestion was run on a 1% agarose gel with an uncut control to confirm linearization and the digested plasmid extracted using a QIAquick Gel Extraction Kit (Qiagen).

The synthesized hDHHC pooled oligo array was amplified by PCR to append homology arms using the following primers and conditions:

DHHCarray_F :

TAACTTGAAAGTATTTCGATTTCTTGGCTTTATATATCTTGTGGAAAGGACGAAAC

ACCG

DHHCarray_R:

TTGTCTCAAGATCTAGTTGATATCGGATCCACGCCAAGCTT

-1µL mDHHC oligo array
-2.5µL DHHCarray_F (10µM)
-2.5µL DHHCarray_R (10µM)
-25µL Phusion Flash High-Fidelity PCR Master Mix (ThermoFisher #F548S)
-19µL ddH_2O

Thermocycler conditions:

1. 98°C 1min.
2. 98°C 1s.
3. 59°C 5s.
4. 72°C 10s. ➔ repeat steps 2-4 for 32 cycles
5. 72°C 2min.
6. 4°C ∞

The PCR product was confirmed by running the reaction on a 1% agarose gel (≈210 bp amplicon) and extracted with a QIAquick Gel Extraction Kit.

The PCR products were then cloned into linearized pRC11 by Gibson assembly according to the following procedure:

-330ng Esp3I-digested pRC11
-50ng PCR-amplified mDHHC library
-10µL 2x Gibson Assembly Master Mix (New England Biolabs #E2611S)
-ddH_2O to 20µL

Thermocycler conditions: 50°C for 60 minutes.

3.3.5 Cpf1 Library Amplification

Of the Gibson reaction product, 2µl was added in separate transformations to Endura Electrocompetent Cells (Lucigen) and electroporated according to manufacturers instructions. After 1hr. of incubation in recovery medium (Lucigen), 10µl of culture was removed, diluted 1:100 and 1:10000, and plated on ampicillin agar to assess transformation efficiency. The remaining transformed cells were streaked on 15cm LB agar plates containing 100µg/ml ampicillin (2x per transformation) and cultured overnight at 37°C. Plasmid DNA was then harvested using a EndoFree Plasmid Maxi Kit (Qiagen) according to the manufacturers instructions. Successful cloning of the sgRNA arrays into the parent vector was confirmed by NotI/XhoI digest against the parent vector.

3.3.6 Viral production and quantification

HEK293T cells were plated at 60-80% confluence in 15cm Biolite culture dishes (Thermo Scientific) and incubated for 24hrs prior to transfecting with pRC49_mDHHC or pRC11_hDHHC library plasmid (8µg), psPAX2 (4.8µg) and PMDG.2 (3.2µg) with Lipofectamine 3000 (Thermo) transfection reagent in Opti-MEM following the manufactures instructions . The cells were incubated for 6hrs. before aspirating and replacing with fresh DMEM culture medium. The medium was then harvested at 24 and 48hrs. post-transfection, sterile-filtered, and stored as l mL aliquots. Viral titer was determined by plating HAP1 cells at

10000 cells/well on 96-well culture plates containing 10-fold serially-diluted TKO viral medium. After 24hrs., the culture medium was replaced with fresh IMDM medium containing 1μg/ml puromycin and cultured another 24hrs. Cell viability was then assessed using WST-1 cell proliferation assay (Roche) and measured cell viability used to determine viral titer.

3.3.7 Generation of Cpf1-Expressing Cell Lines

HEK293T cells were passaged and plated at 30-40% confluence on 6cm Biolite culture dishes and incubated for 24hrs. prior to transfecting with pCLHCX-AsCpf1 (1250ng) and pCL-Ampho (1250ng) with 7.5μl of Fugene HD transfection reagent according to the manufacturer's instructions. The cells were incubated for 24hrs. before aspirating and replacing the medium and incubating for an additional 24hrs. The medium was then harvested at 48 and 72hrs. following transfection and stored at -80°C. Cpf1-expressing HAP1 cell lines were derived from parent cultures by reverse transduction with AsCpf1 retrovirus and selecting for positive clones by incubation with DMEM medium containing 200μg/ml Hygromycin for 72hrs.

3.3.8 Western Blotting

Cells were pelleted by centrifugation, resuspended in 1x DPBS and lysed by sonication. Total protein content was quantified by Lowry assay and 25μg of total protein diluted to final volume of 20μl with DPBS. 1x Lamelli Buffer + SDS was added to the lysates and boiled for 15min at 95°C. Lysates were then vortexed, centrifuged, run on 7% SDS tris-acetate gels, and transferred at 40mV for 12hrs. Transfer membranes were blocked for 1hr with Odyssey Blocking Buffer (Licor) and rocked overnight in OBB + 6x-His Tag Mouse mAb (1:1000) (Invitrogen). The blots were then washed with 1xTBST and incubated in anti-mouse stable peroxidase-conjugated goat

antibody (Invitrogen) for 1hr. at room temperature. The blots were then washed again, treated with SuperSignal™ West Femto Chemiluminescent substrate (Thermo Scientific) and imaged with an Azure c600 gel imager (Azure Biosystems).

3.3.9 hDHHC CRISPR-Cpf1 Duplex Knockout Screen

Cpf1-expressing HAP1 cultures were expanded and transduced with pRC11_hDHHC lentivirus at 300-fold library coverage and an MOI of ≈0.2. After 24hrs incubation, the cell medium was aspirated and replaced with medium containing 1μg/ml puromycin and selection allowed to proceed over 48hrs or until cells reached confluence. All selected cells were then collected and subdivided over 3x replicate 15cm culture dishes at 300-fold coverage and a 300-fold sample retained for each replicate as a screen base control and denoted T_0. The cultures were maintained for 3 weeks under puromycin and hygromycin selection, passaging every 3-4 days to prevent the cultures from reaching confluence and maintaining them at respective initial library coverage. After 18 days (approximately 5 passages), the cultures were collected and sampled at 300-fold coverage(T_{18}). The T_0 and T_{18} time point replicates for each sample were stored at -80°C.

3.3.10 Extraction of Genomic DNA

Cell pellets were thawed and resuspended in 200μl of 1 x DPBS. The genomic DNA was then extracted using DNeasy Blood and Tissue Kit (Qiagen) following the manufacturer's instructions. The total recovered gDNA was quantified using a Tecan NanoQuant plate.

3.3.11 Illumina Library Preparation and Sequencing

crRNAs were recovered from genomic DNA and prepared for Illumina sequencing by nested PCR using methods and primers adapted from Chen et al.

First (outer) PCR:

Genomic DNA :	3µg
$MgCl_2$ (50mM) :	1µL
dNTP Mix (10mM):	1µL
DHHC_vrd1_F (10µM)	2.5µL
DHHC_vrd1_R (10µM)	2.5µL
Herculase Buffer:	10µL
Herculase II Fusion Polymerase (Agilent):	0.5µL
diH_2O:	to 50µL

Thermocycler conditions:

1. 98°C 3min.
2. 98°C 20s.
3. 52°C 30s.
4. 72°C 30s. ➔ repeat steps 2-4 for 22 cycles
5. 72°C 3min.
6. 4°C ∞

5x replicate reactions were conducted for each sample in the screen to ensure sufficient coverage. All sample replicates were pooled following completion of the first PCR.

Second (Barcode) PCR:

Pooled outer PCR product:	4µL
$MgCl_2$ (50mM) :	1µL
dNTP Mix (10mM):	1µL
DHHC Index Adaptor F (10µM)	2.5µL

DHHC Index Adaptor R (10μM)	2.5μL
Herculase Buffer:	10μL
Herculase II Fusion Polymerase (Agilent):	0.5μL
diH_2O:	to 50μL

Thermocycler conditions:

1. 98°C 3min.
2. 98°C 20s.
3. 52°C 30s.
4. 72°C 30s. ➔ repeat steps 2-4 for 16 cycles
5. 72°C 3min.
6. 4°C ∞

All barcode PCR reactions were conducted in duplicate, pooled upon completion and concentrated by spin column purification (Qiagen). The concentrated PCR products were resolved on 1% agarose gels and ≈400bp barcoded products excised and gel purified using a QIAquick gel extraction kit (Qiagen). The HAP1 libraries were sequenced using an Illumina MiSeq with 300-cycle (V2), paired-end settings.

3.3.12 Data Analysis

The raw fastq data was uploaded to the Galaxy server and all subsequent processing conducted therein. In brief, forward and reverse raw fastq files were collapsed into single datasets according to their respective i7 barcodes and subsequently split by their i5 forward barcodes into their respective individual samples. The reads were then filtered by size to remove all reads below 150bp and trimmed to recover the unique 10nt barcode. All trimmed read files were then filtered to remove any reads containing bases with a quality score less than 15. The processed reads were then aligned to all barcode sequences present in the mDHHC or hDHHC libraries using Bowtie2 with default settings. Read counts were then extracted using SAMtools.

3.4 Results

3.4.1 Cloning of hDHHC Cpf1 Library

The hDHHC CRISPR-Cpf1 library was cloned into pRC11 (pLenti-U6-DR-crRNA-BsmbI(x2)/EFS-Puro-WPRE) from oligo arrays synthesized by Genescript using the methods described by Chow et al. Library coverage was assessed by 2nd-generation sequencing using the Illumina MiSeq platform. The extracted reads revealed nearly complete coverage of the originally designed library (≈95% overall) with no discrimination between arrays for loss of coverage. (**Figure 19)** The depth of coverage for nearly all crRNA arrays was found to be at or above a LOG_2 (RPM) of 5 and nearly uniform with few over- and under-represented arrays. (**Figure 20)**

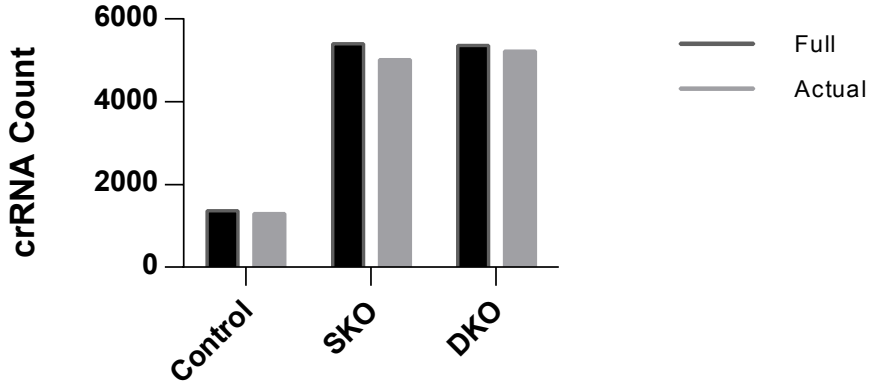


Figure 19. Representation of crRNA arrays by type relative to original library design in hDHHC Cpf1 Library.

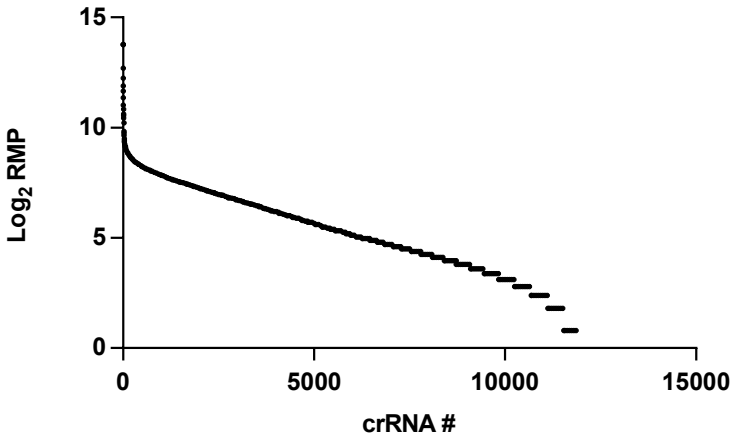


Figure 20. Representation of all crRNAs by count in reads-per-million in the human hDHHC Cpf1 library.

3.4.2 Cloning of mDHHC Cpf1 Library

The mDHHC CRISPR-Cpf1 library was cloned into pRC49 (pLenti-U6- DR-crRNA-BsmbI(x2)-6T/EFS-Puro-2A-Fluc-2A-EGFP_NLS-WPRE) from oligo arrays synthesized by Genescript using identical methods to the previous library. 2nd-generation sequencing of the cloned library revealed a nearly identical level of coverage compared to the hDHHC library (≈93% overall) again, with no discrimination between arrays. (**Figure 21**) The distribution of the sequencing depth across all arrays revealed a greater proportion with representation at or above a LOG_2 (RPM) of 5 than what was observed for the human-specific library. (**Figure 22**) A comparison of the sequencing depth across all represented arrays between the two libraries however revealed little difference in either the average or variance of either. (**Figure 23**)

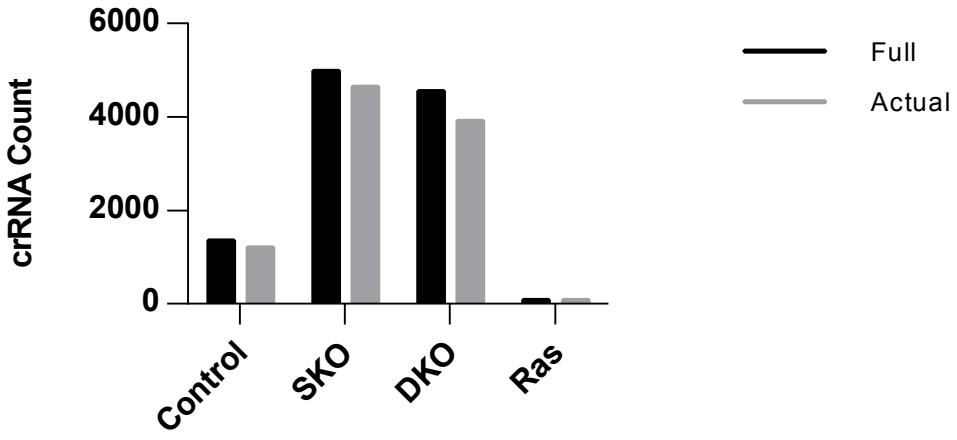


Figure 21. Representation of crRNA arrays by type relative to original library design in mDHHC Cpf1 Library.

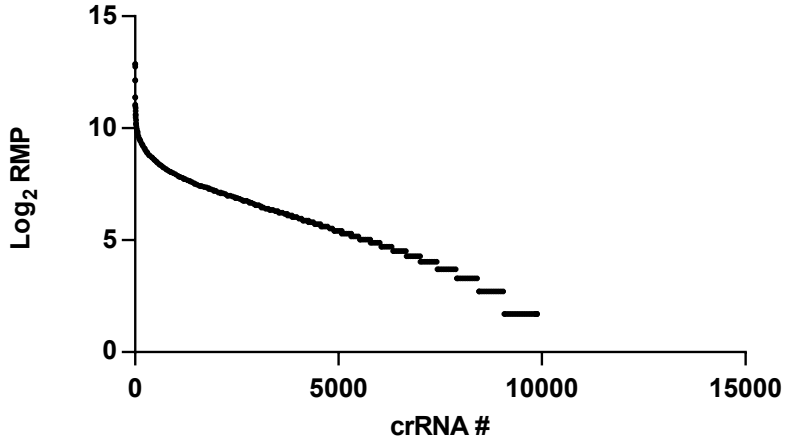


Figure 22. Representation of all crRNAs by count in reads-per-million in the mouse mDHHC Cpf1 library.

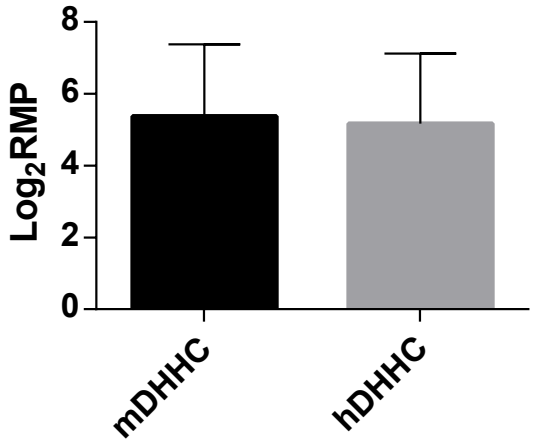


Figure 23. Comparison of average library representation per unique array present in the synthesized hDHHC and mDHHC CRISPR Cpf1 libraries.

3.4.3 HAP1 Screen for DHHC Functional Redundancy

The hDHHC CRISPR library was transduced into HAP1 human haploid cells stably expressing AsCpf1. After 18 days, the cultures were collected, sequenced, and fold changes calculated. Plotting the fold changes of each crRNA array against its statistical p-value revealed a distribution that was little shifted away from the center with the exception of a number of arrays displaying large positive fold changes. When subdivided between the different array types little difference was observed between the single and double knockout arrays although their distributions did appear shifted from that of the nontargeting control arrays. (**Figure 24**)

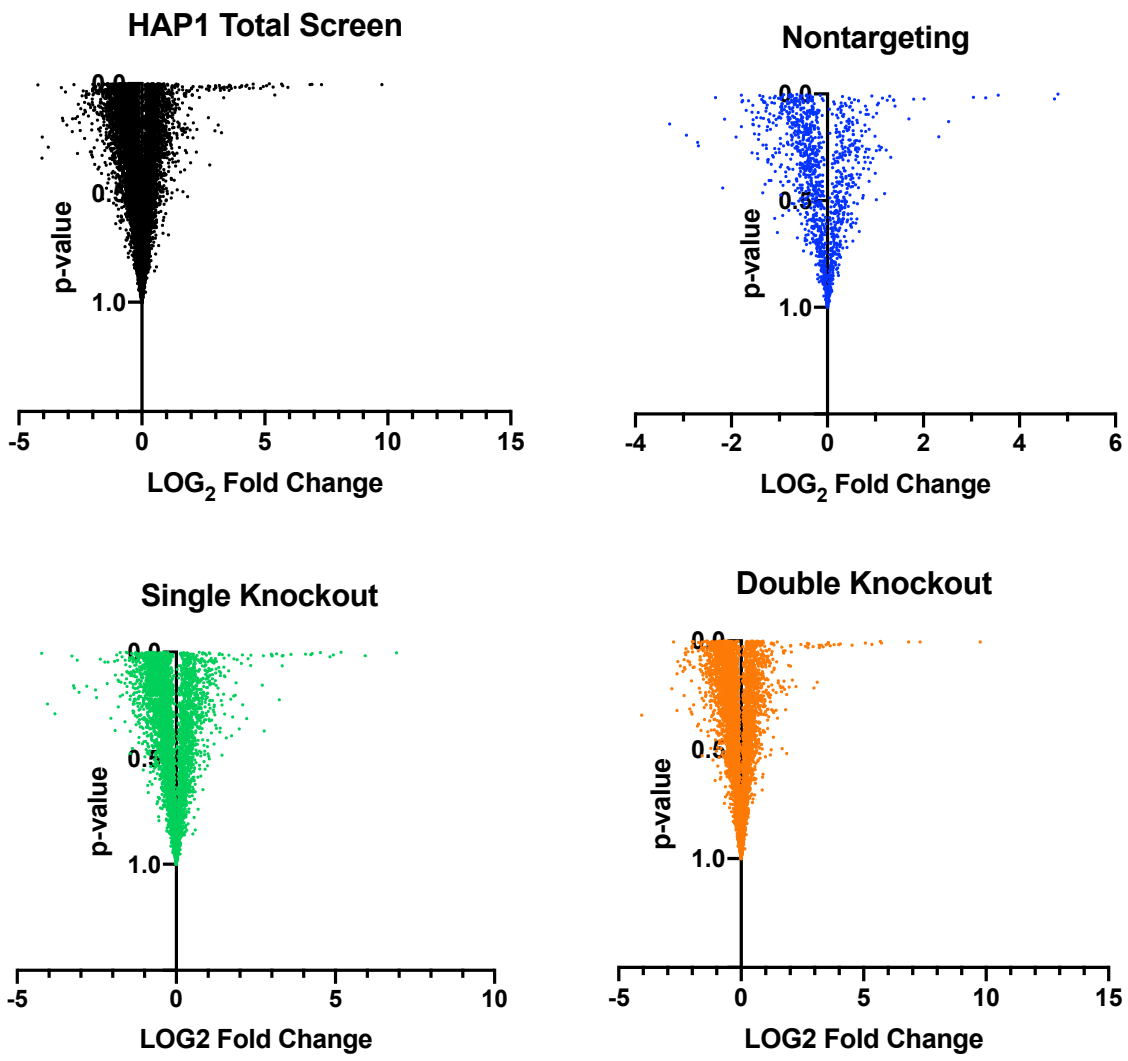


Figure 24. Volcano plots of fold change differences for (clockwise) total screen, nontargeting arrays, single knockout arrays, and double knockout arrays of all recovered crRNA arrays from end of screen (T_{18}) compared to start (T_0).

From these, statistically significant crRNAs exhibiting fold changes of +/- 2 were identified and selected for further analysis. Of the significant single KO arrays only sequences targeted to DHHC18 appeared to be depleted with any frequency. Average fold changes were calculated for all single KO arrays targeted to DHHC18. (**Figure 25**)

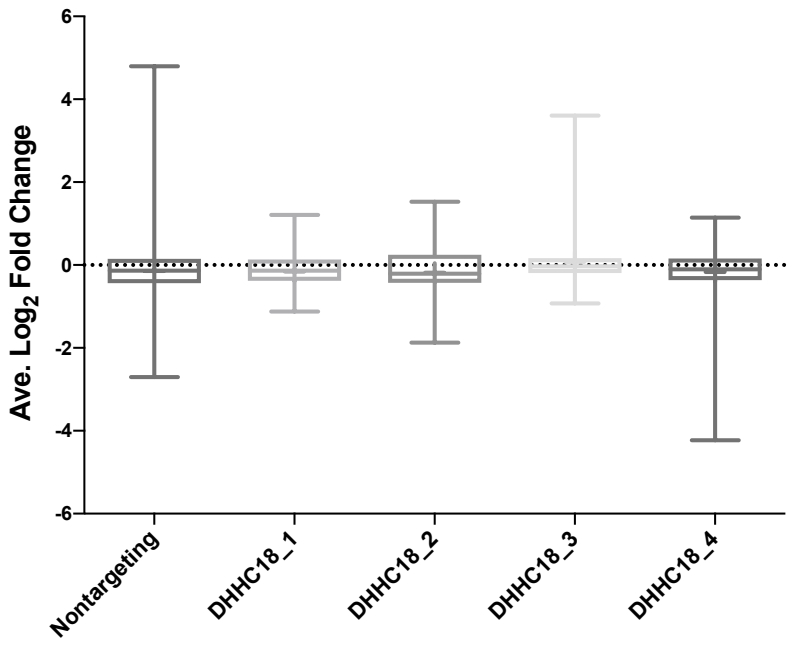


Figure 25. Average $\log_2$ fold change distributions of all single targeting crRNA arrays for DHHC18.

No significant variation was observed for any of DHHC18-targeted arrays compared to the non-targeting controls. With no other significant drop-outs observed for the remaining DHHCs, the SKO arrays for DHHCs exhibiting positive fold changes were then investigated. Plotting the average fold changes for all SKO arrays targeted to each significant DHHC again showed little variation from the non-targeting arrays. (**Figure 26**)

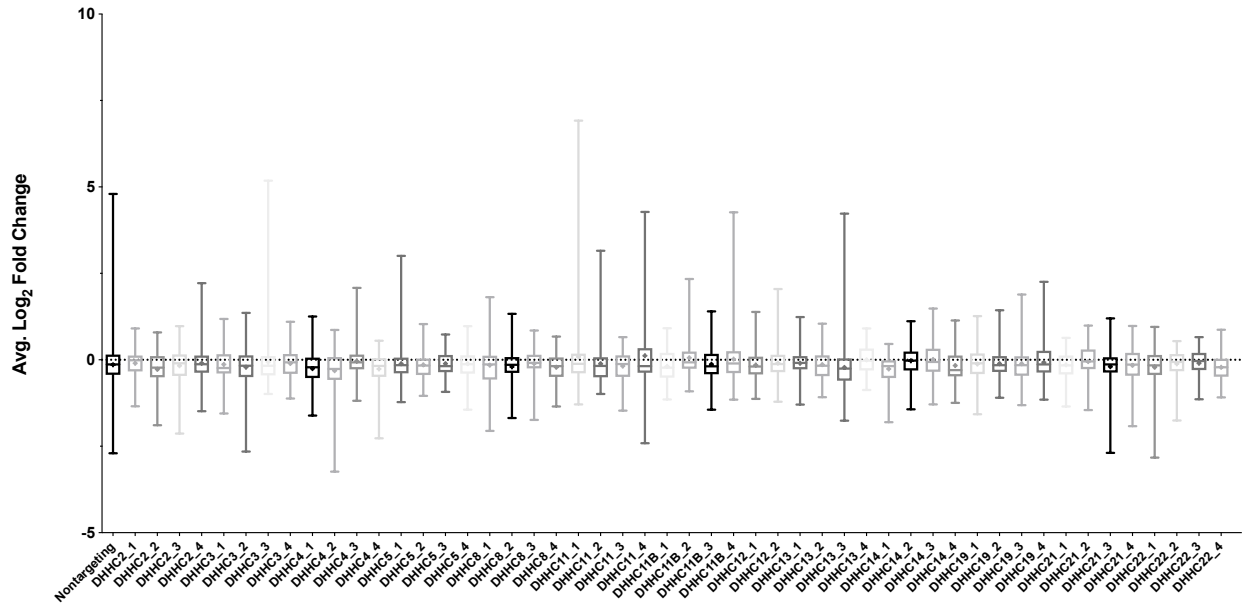
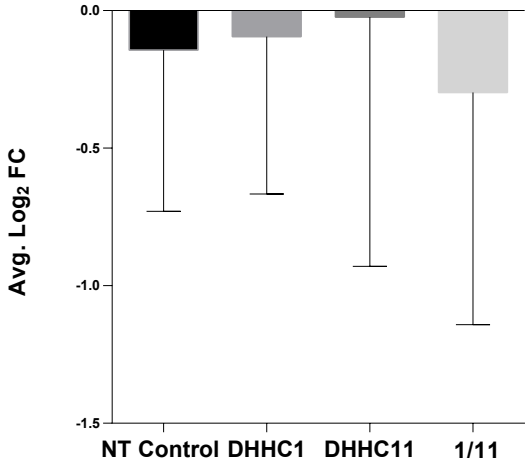
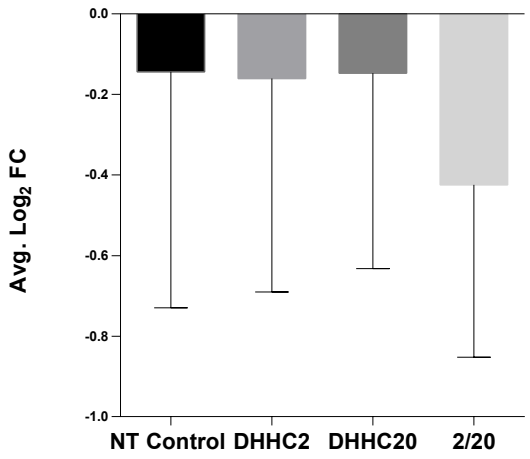
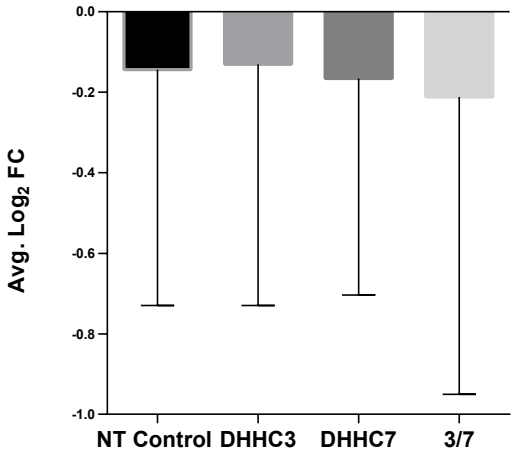


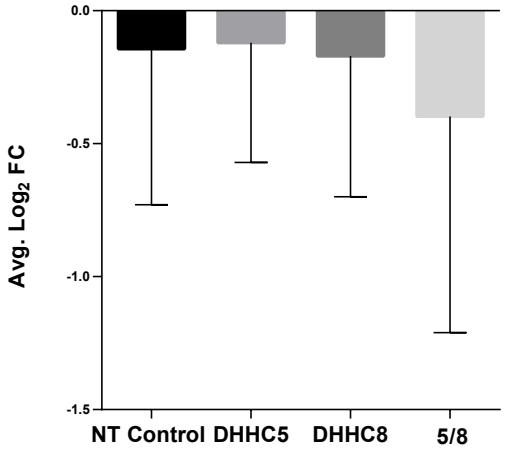
Figure 26. Average $\log_2$ fold change distributions of all single targeting crRNA arrays for all DHHCs exhibiting enrichment over time course of screen.

Only arrays targeted to DHHCs 11 and 11B exhibited average positive fold changes although the low average fold-change and high variability lend doubt to the reliability of this observation. Double knockout arrays were then investigated. Average fold changes for arrays targeted to DHHCs within the same structural subfamilies were compared to their respective single knockout arrays. (**Figure 27**) DHHCs 9 and 14, and 3 and 7 showed no change over their respective single arrays. DHHCs 1 and 11, 13 and 17, and 18 and 19 showed slight average decrease over their single arrays whereas 5 and 8, and 2 and 20 exhibited more notable changes. None of these, however, were found to be significant over non-targeting arrays and so cannot provide evidence of functional redundancy between structurally similar DHHCs.


Avg. Log2 FC
0.0
-0.5
-1.0
-1.5
NT Control
DHHC1
DHHC11
1/11


Avg. Log2 FC
0.0
-0.2
-0.4
-0.6
-0.8
-1.0
NT Control
DHHC2
DHHC20
2/20


Avg. Log2 FC
0.0
-0.2
-0.4
-0.6
-0.8
-1.0
NT Control
DHHC3
DHHC7
3/7


Avg. Log2 FC
0.0
-0.5
-1.0
-1.5
NT Control
DHHC5
DHHC8
5/8

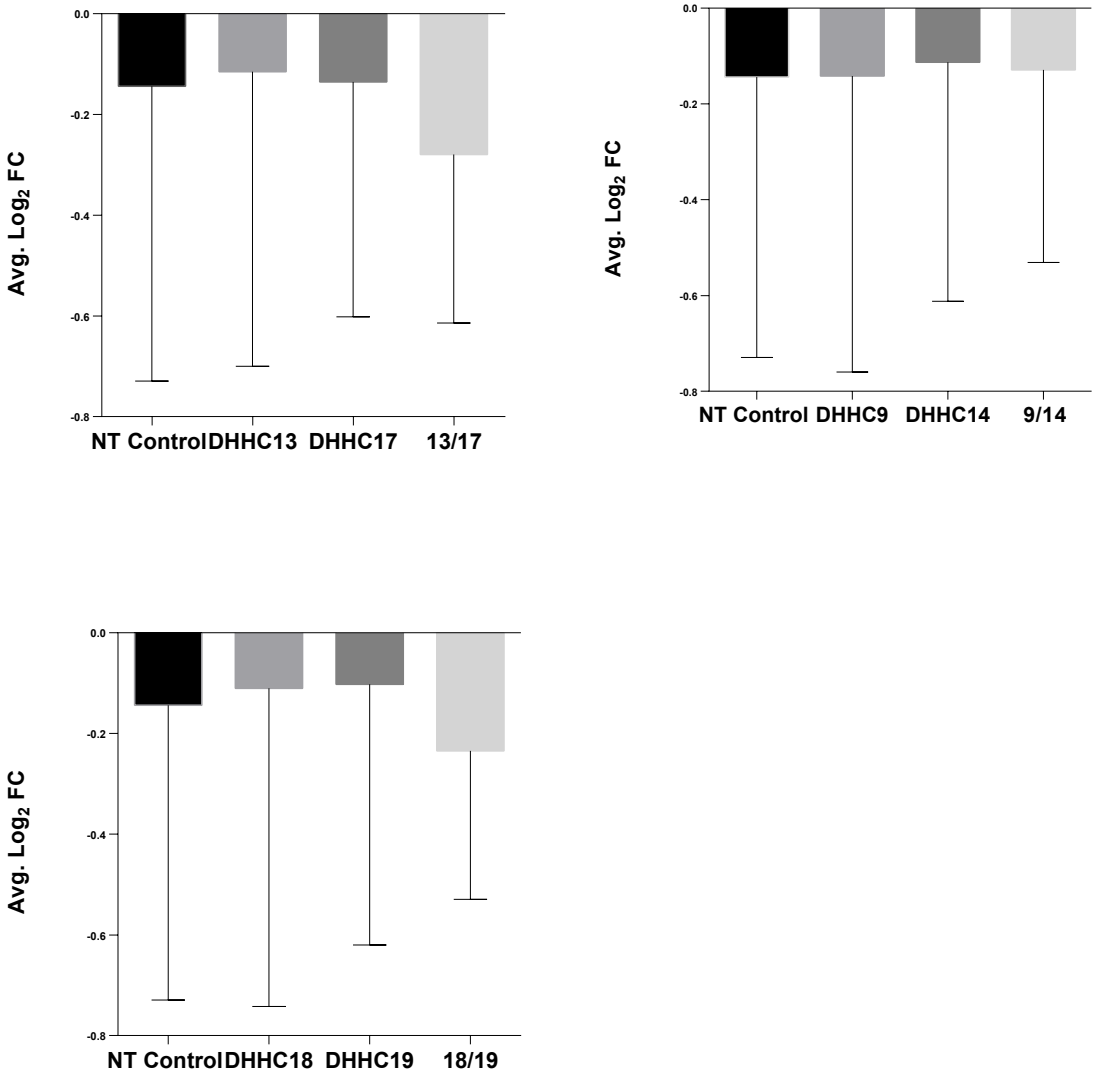


Figure 27. Average $\log_2$ fold change of all single targeting crRNA arrays for all DHHC subfamilies compared to duplex arrays.

Finally, to determine if these insignificant changes were due to inefficiency in Cpf1 or to an inherent lack of genetic sensitivity in the HAP1 cell model to DHHC KO, the single knockout arrays targeted to previously reported essential genes were investigated. The distribution of arrays targeted to essential genes was found to be little different to that of the nontargeting controls when compared. (**Figure 28**) Closer investigation of these groups once again revealed

little change between the single knockout arrays and the non-targeting controls. (**Figure 29**) This suggested that flaws in the design of the library or a lack of efficiency in AsCpf1 prevented the detection of a growth phenotype.

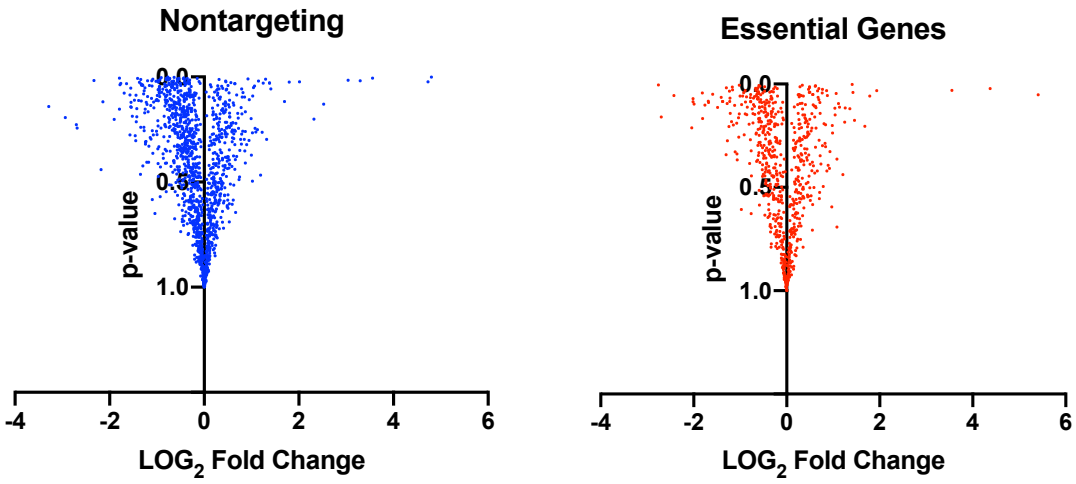


Figure 28. Volcano plots of fold change differences for nontargeting and essential gene arrays, from end of screen (T_{18}) compared to start (T_0).

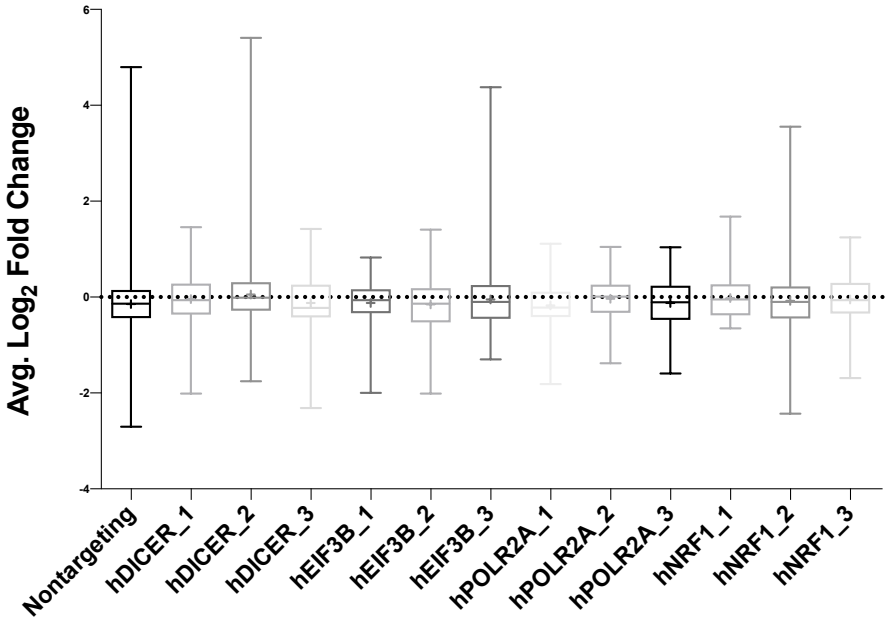


Figure 29. Average $\log_2$ fold change of all single targeting crRNA arrays for previously validated essential genes.

3.5 Discussion

Cpf1 offers many advantages over Cas9 for performing pooled genetic screens. Unlike Cas9, Cpf1 is capable of processing its own mature crRNAs from pre-crRNAs and does not require a tracrRNA to form an active complex capable of cutting genomic DNA. This allows for multiple crRNAs to be expressed in tandem from a single promoter and opens the possibility of creating multiple simultaneous targeted gene knockouts. This should, in principle make Cpf1 uniquely suited to combinatorial pooled genomic screens. Herein, we designed a duplex knockout library with sgRNAs targeted to all DHHCs expressed in the human genome to assess the prevalence of functional redundancy within this gene family. Prior to performing these screens, the efficiency of Cpf1 was assessed by western blot in MEF cells using single sgRNAs targeted to the thioesterase APT1. For each of the three sgRNAs tested a small but noticeable change was observed in the amount of APT1 present over sgRNAs that lacked a targeting sequence and this effect became more pronounced when multiple sgRNAs were employed. When Cpf1 was used in a pooled screen of HAP1 cells however, no significant change was observed for any DHHC for either single or double knockout arrays. There are three possible factors that may have contributed to this negative result. Although DHHCs are known to palmitoylate many proteins involved in cellular growth it is possible that the knockout of one or more DHHCs may not be sufficient to create a noticeable deficiency in the growth of the affected cell. While this is certainly conceivable, and every effort should be made when performing genomic screens to use a strictly controlled genetic model, it is unlikely as the same absence of a phenotype was observed for sgRNAs targeted to essential genes determined from screens performed across multiple cell models. Although it has been extensively investigated in Cas9, little is known about the principles that determine the efficacy of sgRNAs for Cpf1. Here, we selected 3-4 sgRNAs for

each targeted gene based solely on their position in the coding sequence and their off-target score (likelihood of cutting at locations other than the intended target). While unlikely, it is possible that the sequences selected lacked sufficient activity toward their targets to affect a noticeable phenotype. A systematic investigation into the rules that govern Cpf1 sgRNA activity is certainly needed to ensure efficient editing of target genes. The final possibility is that Cpf1 lacks the necessary efficiency to generate sufficient gene knockouts to be observed in a genetic screen. Recently, Liu et al.[127] performed a systematic study of AsCpf1 pooled genetic screens. Here, they found that the commonly used AsCpf1 variant (the same used here) consistently failed to generate separation between validated essential and non-essential genes in a pooled genetic screen despite showing no deficiency in generating indels from individual constructs. This effect was remedied by employing an AsCpf1 variant containing a 3x MYC NLS, revealing that the previously noted inefficiency of Cpf1 compared to Cas9 may be due in part to inefficient nuclear translocation. Furthermore, they found that the knockout efficiency improved when constructs containing multiple sgRNAs to a particular gene were used. While the constructs here contained sgRNAs targeted to different genes this might explain, in part, why double knockout arrays seemed to outperform their single knockout counterparts although even these differences were, at best, marginal. It would be of interest to repeat these screens using this optimized Cpf1 to see if this remedies the inefficiencies observed here. Presuming that the failure of our screens derives entirely from insufficient activity by Cpf1 it would be reasonable to expect that a repetition of these would yield a significant depletion phenotype at least for crRNAs targeted to essential genes. As a preliminary to these repeat screens, expression vectors identical to those contained in the libraries could be individually cloned and introduced into their respective cell models to assess them for gene-dependent growth depletion and provide assurances that this new Cpf1

variant is operating as expected. These vectors would contain the singly and doubly targeted arrays to essential genes or Ras isoforms present as controls in the hDHHC and mDHHC libraries respectively and would provide unambiguous validation of Cpf1 as well as evidence of the impact of double verses single crRNA on the knockout efficiency of the gene targeted. Should the growth assays for individual essential genes and the HAP1 screen provide good evidence of proper screen function we could then proceed to our original experimental aims and perform our screens for functional redundancy in Rasless MEFs using the mDHHC library. Should these preliminary experiments show otherwise and Cpf1 3x Myc NLS fail to sufficiently correct these problems, a redesign of the crRNA libraries would then need to be undertaken. As previously stated, Liu et al.[127] found that the sensitivity of Cpf1 CRISPR screens only approached that of Cas9 when three crRNAs targeted to the same gene were expressed in tandem from a single expression vector. This suggests that randomizing single crRNAs targeting different genes into duplex arrays may be insufficient to generate a detectable phenotype and present an additional obstacle to proper screen function. To amend this issue, the libraries will be redesigned using randomized gene blocks (2-3 crRNAs targeted to a single gene) in place of single crRNAs, resulting in the expression of 4-6 crRNAs directed against two different genes for each expression vector. (**Figure 30**) The differences in observed Cpf1 activity with increasing numbers of crRNAs per gene block should be first evaluated using single expression vectors containing crRNAs directed against select essential genes prior undertaking a complete redesign of CRISPR libraries.

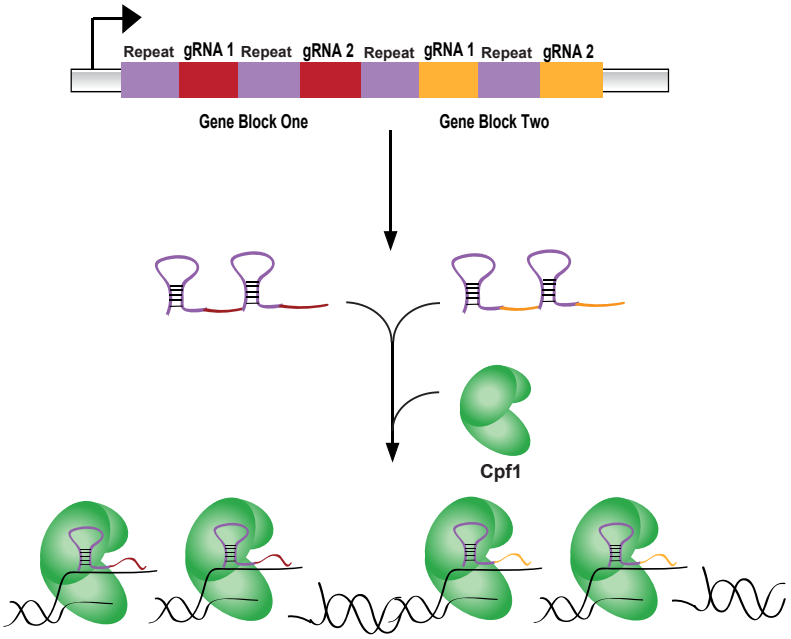


Figure 30. An alternative approach to the design of duplex CRISPR-Cpf1 knockout libraries involves the use of randomized blocks of crRNAs targeted to the same gene in place of single crRNAs.

Once designed and assembled, these improved CRISPR-Cpf1 knockout libraries will then be employed in our human and, ultimately, MEF cell models to assess the extent of functional redundancy in DHHC palmitoyltransferases. An alternative approach that may be used in support or in place of these Cpf1 screens as necessary would make use of selective enrichment of palmitoylated proteins and LC-MS proteomics to assess the substrate profiles of individual DHHC PATs and determine where they are unique and where they overlap. The overexpression and/or knockout of single DHHCs can be undertaken in metabolically labeled cell culture[128] and the isotopically-labeled proteomes enriched for palmitoylated proteins using acyl resin-assisted capture (acyl-RAC).[129] (**Figure 31)** A panel of palmitoylated substrates could then be determined for each palmitoyltransferase by quantifying changes in palmitoylation relative to catalytic dead (overexpression) or wild-type (knockout) controls. If redundancy exists between PATs as expected, it would appear in the form of shared substrates. This approach would provide a

straightforward answer to the question of functional redundancy and, furthermore, would provide a more comprehensive analysis of the native substrates for every DHHC investigated.

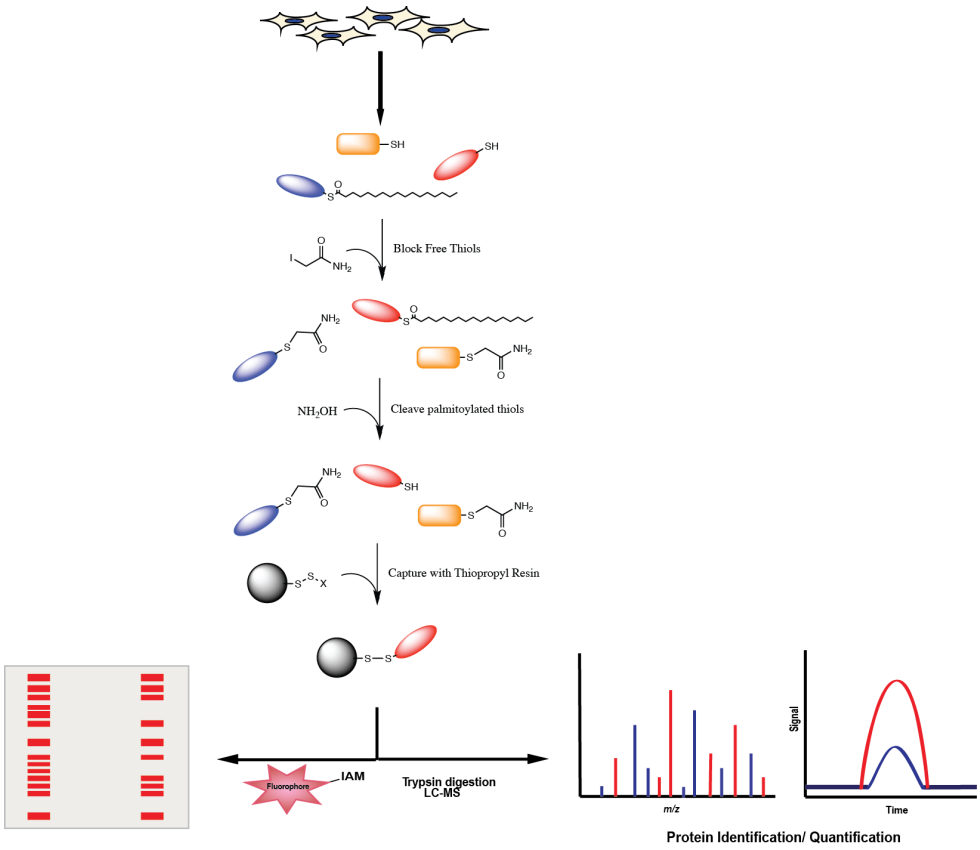


Figure 31. Acyl resin-assisted capture may be used to enrich for palmitoylated proteins for further proteomic study. Free thiols are first blocked by covalent modification after lysing the cells. Palmitoylated thiols are then cleaved with hydroxylamine and the now-free thiols captured with a thiol resin. The captured proteins may then be collected and prepared for proteomic analysis.

3.6 Conclusions

Questions surrounding the existence of functional redundancy between DHHC family palmitoyltransferases have yet to be systematically addressed. Herein, we attempted to investigate the extent of functional redundancy between DHHC family palmitoyltransferases by performing multiplex CRISPR-Cpf1 genetic screens. Duplex libraries targeting all know DHHC proteins in mice and humans respectively were successfully designed and cloned into lentiviral

expression vectors. Initial attempts to use these libraries in pooled screens however failed to uncover a significant depletion phenotype even in those arrays targeted to essential genes that had been previously validated. Recently published literature suggests that the suboptimal performance of these screens may be due either to insufficient activity of commonly used Cpf1 variants or an insufficient number of crRNAs included into each expression vector. Newer Cpf1 variants have since been engineered to address issues surrounding Cfp1 activity in pooled screens and should be tested with the CRISPR libraries designed here to assess whether a complete redesign is necessary to achieve the intended experimental aims. Once proper screen function is confirmed after making the necessary changes, screens should be performed in appropriate human and mouse cell models supported by DHHC overexpression/knockout LC-MS proteomics to provide a comprehensive assessment of PAT functional redundancy, substrate specificity, and the site-specific palmitoylation of protein substrates.

Chapter 4

Conclusions and Future Directions

4.1 Overview of Experimental Outcomes

Herein, we attempted to use CRISPR pooled genetic screens to identify the necessary components of the cellular oxidative stress response and investigate the question of functional redundancy between structurally similar DHHC-family acyltransferases. The first of these was accomplished in part using the Cas9 TkOv1 pooled CRISPR library although the small number of identified hits and the lack of known functional uniformity between them suggests that many of these may be false positives created by insufficient genotoxic stress of the hydrogen peroxide treatment used. The second question, unfortunately, remains completely unresolved by our efforts, accomplished here by employing a custom Cpf1 duplex CRISPR library. This is most likely due to inefficiency in the activity of AsCpf1 to induce sufficient indels necessary to achieve a discernable phenotype in the context of a pooled screen. This effect was noted by Liu et al.[127] in their recent comprehensive study of Cpf1-CRISPR pooled screens and was attributed to inefficient nuclear localization of commonly employed AsCpf1 variants. Indeed, they found that this inefficiency was almost entirely corrected by employing a novel 3x Myc-tagged AsCpf1 variant such that it's performance in pooled knockout screens approached that of Cas9. Although our primary research aims were not accomplished using these approaches they do, however, provide lessons in the design and use of CRISPR systems for both single and genome-wide knockout that may be of use to others intending to use these methods. Herein, we will summarize

what may be considered best practices for using CRISPR methods drawing on the current literature and supplemented by our own experiences.

4.2 Considerations for Optimal crRNA Design

Efficient gene editing using CRISPR requires that the sgRNAs used consistently guide the nuclease complex to its gene target and induce cutting. Ideal sgRNAs should induce efficient cleavage at their target location while not creating off-target cleavages at other regions within the genome. Several studies into the factors that determine sgRNA efficiency have been conducted and from these a series of rules have been determined for the design of sgRNAs. For Cas9 sgRNAs, the composition of the targeting sequence and the sequence of the PAM have been shown to have the greatest influence on activity. Of the 20bp targeting sgRNA, the 8-14bp region proximal to the PAM have been found to be essential to activity with even single mismatches resulting in a significant reduction in activity at the intended target site and a concurrent increase in off-target activity; the distal regions, by contrast, had much less influence on activity.[128] The activity of sgRNAs was further influenced by nucleotide composition, gRNA location within the protein, and DNA melting temperature.[129] The optimal PAM sequence has been repeatedly show to be 5'-NGG with sgRNAs possessing a C at the first position exhibiting the highest activity those possessing a T the lowest.[130] These design factors have been incorporated into a number of algorithms to confidently identify sgRNAs with high activity.[131] For the selection of sgRNAs, it would be advisable to select previously validated sequences, however this is only possible for a small subset of genes. In these situations, it is recommended that these validated scoring algorithms be used to identify sgRNAs that possess both high activity and exhibit low off-target activity. Furthermore, it has been shown that the region within a gene targeted by the sgRNA has little effect on its activity with the exception of the C-

terminus.[132] This encourages the selection of sgRNAs spaced across the length of target gene to improve the number of options for selecting sgRNAs with high activity and specificity. To date, no systematic study of Cpf1 sgRNA activity comparable to Cas9 has been conducted and so it is uncertain what differences, if any, exist between these two systems. Elucidation of these Cpf1-specific design principles will be necessary to maximizing the advantages offered by this system. To mitigate potential issues with the selection of Cpf1 sgRNAs with low activity multiple sgRNAs should be selected for each target gene, spaced across the length of the coding sequence.

4.3 Considerations for Optimal CRISPR Screen Design

Effective genomic screening with CRISPR requires that a balance be maintained between library complexity and the scale of the proposed screen. Early CRISPR libraries required large numbers of sgRNAs be employed for each gene to increase sensitivity by circumventing unknown sources of inefficiency. When this strategy was applied for whole-genome screens, the resulting library complexity required high numbers of cultured cells to maintained sufficient library coverage for the screen to function properly. This imposed technical limitations on the scale of the screens and the cell models that could be employed in them and therefore, the questions that could be answered using CRISPR screens. A better understanding of the principles that govern sgRNA activity permitted a reduction in the number of sgRNAs necessary to effectively perform whole genome screens. This was aptly demonstrated by Hart et al. in the design of their TKOv3 CRISPR library where the library size was reduced from 176,500 to 71,090 sgRNAs over their TKOv1 library while significantly improving library performance.[133] Library complexity can be further reduced when the questions being investigated necessitate that only single class or subset

of genes be employed in the screen. This has been done to create Cas9 libraries for metabolic genes, kinases, and DNA-binding domain proteins among others at a fraction of the library size required for whole-genome screening.[134,135] When designing pooled CRISPR screens for Cas9, it is therefore important to carefully consider the question the screen is intended to answer so as to appropriately design the library with the minimum requisite complexity for the scope of the screen to be performed. The unique ability of Cpf1 to process multiple crRNAs from single arrays provides for improved flexibility in screen design over Cas9. This feature not only permits screening of combinatorial interactions between different genes but can also be used to further reduce the complexity of whole-genome libraries without sacrificing sensitivity. This was aptly demonstrated by Liu et al. in their design of a multiplex AsCpf1 library where 2,061 distinct guides were assembled into 687 library constructs by combining all guides targeted to a particular gene into a single three guide array.[127] This library showed comparable performance to an identical mono-cistronic Cas9 library of 2,061 constructs. Interestingly, these two libraries significantly outperformed an identical mono-cistronic AsCpf1 library suggesting that AsCpf1 requires multiple crRNAs to be expressed simultaneously to induce efficient gene editing when used in pooled screens. This raises questions about the sensitivity of AsCpf1 when used to screen combinatorial interactions between different genes; it may be necessary to pair 2-3 guide arrays targeted to each gene in order to reproduce the efficiency demonstrated here for the screening of single genes.

4.4 General Guidelines for CRISPR Genomic Screen Design

The importance of careful library design cannot be understated when using CRISPR pooled screens to investigate biological phenomena. Our experiences working with CRISPR pooled

screens have demonstrated that equal consideration should be given to the implementation of the screen itself. Here, particular consideration is given to cell model used in the screen. In both our Cas9 and Cpf1 screens we employed a nearly haploid human HAP1 cell line. This cell line grows to high density and was aptly suited to whole genome knockout screens particularly those performed with the larger first generation CRISPR libraries. Although it was not the primary cause of our Cpf1 screens failure, it must be pointed out that the DHHC-family proteins targeted by the screen may not have been sufficiently important to cell survival to induce a detectable phenotype in this cell line. A better approach would have been to employ a cell line with a carefully controlled growth dependency such as that present in the Rasless MEF cell lines that were briefly used herein. Similarly, when the genetic components of a particular stressor are being investigated compared to an untreated population it is critical that the treatment exert its effects through a defined mechanism of action. This is aptly demonstrated by the early CRISPR KO screens performed using cytotoxic drugs such as vemurafenib.[136] These stressors operate through a defined mechanism and so yielded unambiguous phenotypes that were easily detected by CRISPR knockout screens. In our own screens, we employed a hydrogen peroxide solution applied by daily bolus dose as a source of oxidative stress. This was necessary for want of better methods for creating conditions of oxidative stress in cell culture. This may not have been sufficient however in specificity or duration to parse out a defined genetic phenotype. Hydrogen peroxide induces indiscriminate cell damage and the dosing method used here may have simply induced necrotic cell death independent of the underlying genetic background leading to an abundance of false positives. A better approach would have been knockout some essential component of redox homeostasis such as glutathione cycling so as to induce a constitutive state of oxidative stress. In this genetic model, subsequent knockout of other essential components of

redox homeostasis by CRISPR would be unsustainable and lead to a lethal phenotype. These considerations should be taken into account should this or similarly indiscriminate stressors be used in the context of a pooled genomic screen.

4.5 Conclusions

Although our experiences using CRISPR genetic knockout screens did not yield answers to the questions that we intended to investigate there are abundant lessons to be drawn from them on the proper implementation and limitations of CRISPR methods. It is our hope that the above discussion of CRISPR screen design and implementation will help guide the work of others to a successful outcome. These methods have been, from the time of their discovery to the present, in a state of constant development. They will continue to be refined and applied to new questions. They will certainly lead to new insights beyond the reach of the methods that preceded them and push the boundaries of our understanding of life itself.

References

1. Doudna, JA., Charpentier, E. The new frontier of genome engineering with CRISPR-Cas9. *Science*. **2014**. *346* (6213).

2. Smithies, O., Gregg, RG., Boggs, SS., Koralewski, MA., Kucherlapati, RS. Insertion of DNA sequences into the human chromosomal β-globin locus by homologous recombination. *Nature*. **1985**. *317* (19), 230-234.

3. Thomas, KR., Folger, KR., Capecchi, MR. High Frequency Targeting of Genes to Specific Sites in the Mammalian Genome. *Cell*. **1989**. *44*, 419-428.

4. Strobel, SA., Dervan, PB. Site-specific Cleavage of a Yeast Chromosome by Oligonucleotide-Directed Triple-Helix Formation. *Science*. **1990**. *249* (4964), 73-75.

5. Wang, G., Levy, DD., Seidman, MM., Glazer, PM. Targeted Mutagenesis in Mammalian Cells Mediated by Intracellular Triple Helix Formation. *Mol Cell Biol.* **1995**. *15* (3), 1759-68.

6. Faruqui, AF., Seidman, MM., Segal, DJ., Carroll, D., Glazer, PM. Recombination Induced by Triple-Helix-Targeted DNA Damage in Mammalian Cells. *Mol Cell Biol.* **1996**. *16* (12), 6820-6828.

7. Reza, F., Glazer, PM. Therapeutic Genome Mutagenesis Using Synthetic Donor DNA and Triplex-Forming Molecules. *Methods Mol Biol*. **2015**. *1239*, 39-73.

8. Gupta, SK., Shukla, P. Gene editing for cell engineering: trends and applications. *Crit Rev Biotechnol*. **2017**. *37* (5), 672-684.

9. Kim, YG., Cha, J., Chandrasegaran, S. Hybrid restriction enzymes: Zinc finger fusions to *Fok* I cleavage domain. *P Natl Acad Sci USA*. **1996**. *93*, 1156-1160.

10. Christian, M., Cermak, T., Doyle, EL., Schmidt, C., Zhang, F., Hummel, A., Bogdanove, AJ., Voytas, DF. Targeting DNA Double-Stranded Breaks with TAL Effector Nucleases. *Genetics*. **2010**. *186* (2), 757-761.

11. Miller, JC., Holmes, MC., Wang, J., Guschin, DY., Lee, YL., Rupniewski, I., Beausejour, CM., Waite, AJ., Wang NS., Kim, KA., Gregory, PD., Pabo, CO., Rebar, EJ. An improved zinc-finger nuclease architecture for highly specific genome editing. *Nat Biotech*. **2007**. *25* (7), 778-785.

12. Urnov, FD., Miller, JC., Lee, YL., Beausejour, CM., Rock, JM., Augustus, S., Jamieson, AC., Porteus, MH., Gregory, PD., Holmes, MC. Highly efficient endogenous human gene correction using designed zinc-finger nucleases. *Nature*. **2005**. *435*, 646-651.

13. Ding, Q., Regan, SN., Xia, Y., Oostrom, LA., Cowan, CA., Musunuru, K. Enhanced Efficiency of Human Pluripotent Stem Cell Genome Editing through Replacing TALENs with CRISPRs. *Cell Stem Cell*. **2013**. *12*, 393-394.

14. Timmons L., Tabara, H., Mello, CC., Fire, AZ. Inducible Systematic RNA Silencing in *Caenorhabditis elegans*. *Mol Biol Cell*. **2003**. *14*, 2972-2983.

15. Fraser, A. Human genes hit the big screen. *Nature*. **2004.** *428*, 375-377.

16. Ngo,VN., Davis, RE., Lamy, L., Yu, X., Zhao, H., Lenz, G., Lam, LT., Dave, S., Yang, L., Powell, J., Staudt, LM. A loss-of-function RNA interference screen for molecular targets in cancer. *Nature*. **2006**. *441*, 106-110.

17. Moffat, J., Grueneberg, DA., Yang, X., Kim SY., Kloepfer, AM., Hinkle G., Piqani, B., Eisenhaure, TM., Luo, B., Grenier, JK., Carpenter, AE., Foo, SY., Stewart, SA., Stockwell, BR., Hacohen, N., Hahn, WC., Lander, ES., Sabatini, DM., Root, DE. A Lentiviral RNAi Library for Human and Mouse Genes Applied to an Arrayed Viral High-Content Screen. *Cell*. **2006**. *124* (6), 1283-1298.

18. Luo, J., Emanuele, MJ., Li, D., Creighton, CJ., Schlabach, MR., Westbrook TF., Wong, KK., Elledge, SJ. A Genome-wide RNAi Screen Identifies Multiple Synthetic Lethal Interactions with the Ras Oncogene. *Cell*. **2009**. *137* (5), 835-848.

19. Menddes-Pereira, AM., Sims, D., Dexter, T., Fenwick, K., Assiotis, I., Kozarewa, I., Mitsopoulos, C., Hakas, J., Zvelebil, M., Lord, CJ., Ashworth, A. Genome-wide functional screen identifies a compendium of genes affecting sensitivity to tamoxifen. *P Natl Acad Sci USA*. **2012**. *109* (8), 2730-2735.

20. Brass, AL., Dykxhoorn, DM., Benita, Y., Yan, N., Engelman, A., Xavier, RJ., Lieberman, J., Elledge, SJ. Identification of Host Proteins Required for HIV Infection Through a Functional Genomic Screen. *Science*. **2008**. *319* (5865), 921-926.

21. Boutros, M., Ahringer, J. The art and design of genetic screens: RNA interference. *Nat Rev Genet*. **2008**. *9*, 554-566.

22. Hartenian, E., Doench, JG. Genetic screens and functional genomics using CRISPR/Cas9 technology. *FEBS*. **2015**. *282*, 1383-1393.

23. Barrangou, R., Fremaux, C., Deveau, H., Richards, M., Boyaval, P., Moineau, S., Romero, D., Horvath, P. CRISPR Provides Acquired Resistance Against Viruses in Prokaryotes. *Science*. **2007**. *315*, 1709-1712.

24. Cong, L., Ran, F., Cox, D., Lin, S., Barretto, R., Habib, N., Hsu, P., Wu, X., Jiang, W., Maraffini, L., Zhang, F. Multiplex Genome Engineering Using CRISPR/Cas Systems. *Science*. **2013**. *339*, 819-823.

25. Jinek, M., Chylinski, K., Fonfara, I., Hauer, M., Doudna, J. Charpentier, E. A Programmable Dual-RNA-Guided DNA Endonuclease in Adaptive Bacterial Immunity. *Science*. **2012**. *337*, 816-821.

26. Shalem, O.; Sanjana, N.; Zhang, F. High-throughput functional genomics using CRISPR-Cas9. *Nat Rev Genet*. **2015.** 1*6*, 299-311.

27. Hwang, W., Fu, Y., Reyon, D., Maeder, M., Tsai, S., Sander, J., Peterson, R., Yeh, J., Joung, J. Efficient genome editing in zebrafish using a CRISPR-Cas system. *Nat Biotechnol.* **2013**. *31*, 227-229.

28. Sternberg, S., Doudna, J. Expanding the Biologist's Toolkit with CRISPR-Cas9. *Mol Cell.* **2015**. *58*, 568-574.

29. Konermann, S., Brigham, M., Trevino, A. Joung, J., Abudayyeh, O., Barcena, C.; Hsu, P., Habib, N., Gootenberg, J., Nishimasu, H., Nureki, O. Zhang, F. Genome-scale transcriptional activation by an engineered CRISPR-Cas9 complex. *Nature.* **2015**. *517*, 583-588.

30. Gilbert, L., Larson, M., Morsut, L., Liu, Z., Brar, G., Torres, S., Stern-Ginossar, N., Brandman, O., Whitehead, E., Doudna, J., Lim, W., Weissman, J., Qi, L. CRISPR-Mediated Modular RNA-Guided Regulation of Transcription in Eukaryotes. *Cell*. **2013**. *154*, 442-451.

31. Gilbert, L., Horlbeck, M., Adamson, B., Villalta, J., Chen, Y., Whitehead, E., Guimaraes, C., Panning, B., Ploegh, H., Bassik, M., Qi, L., Kampmann, M., Weissman, J. Genome-Scale CRISPR-Mediated Control of Gene Repression and Activation. *Cell*. **2014**. *159*, 647-661.

32. Shalem, O., Sanjana, NE., Hartenian, E., Shi, X., Scott, DA., Mikkelsen, TS., Heckl, D., Ebert, BL., Root, DE., Doench, JG., Zhang, F. Genome-Scale CRISPR-Cas9 Knockout Screening in Human Cells. *Science*. **2014**. *343*, 84-87

33. Hart, T., Chandrashekhar, M., Aregger, M., Steinhart, Z., Brown, KR., MacLeod, G., Mis, M., Zimmermann, M., Fradet-Turcotte, A., Sun, S., Mero, P., Dirks, P., Sidhu, S., Roth, FP., Rissland, OS., Durocher, D., Angers, S., Moffat, J. High-Resolution CRISPR Screens Reveal Fitness Genes and Genotype-Specific Cancer Liabilities. *Cell*. **2015**. *163* (6), 1515-1526.

34. Evers, B., Jastrzebski, K., Heijmans, JPM., Grernrum, W., Beijersbergen, RL., Bernards, R. CRISPR knockout screening outperforms shRNA and CRISPRi in identifying essential genes. *Nat Biotechnol.* **2016**. *34* (6), 631-634.

35. Wang, T., Yu, H., Hughes, NW., Liu, B., Kendirli, A., Klein, K., Chen, WW., Lander, ES., Sabatini, DM. Gene Essentiality Profiling Reveals Gene Networks and Synthetic Lethal Interactions with Oncogenic Ras. *Cell*. **2017**. *168* (5), 890-903.

36. Marceau, CD., Puschnik, AS., Majzoub, K., Ooi, YS., Brewer, SM., Fuchs, G., Swaminathan, K., Mata, MA., Elias, JE., Sarnow, P., Carette, JE. Genetic dissection of *Flaviviridae* host factors through genome-scale CRISPR screens. *Nature*. **2016**. *535*, 159-163.

37. Park, RJ., Wang, T., Koundakjian, D., Hultquist, JF., Lamothe-Molina, P., Monel, B., Schumann, K., Yu, H., Krupzcak, KM., Garcia-Beltran, W., Piechocka-Trocha, A., Krogan, NJ., Marson, A., Sabatini, DM., Lander, ES., Hacohen, N., Walker, BD. A genome-wide CRISPR screen identifies a restricted set of HIV host dependency factors. *Nat Genet.* **2017**. *49* (2), 193-203.

38. Jost, M., Chen, Y., Gilbert, LA., Horlbeck, MA., Krenning, L., Menchon, G., Rai, A., Cho, MY., Stern, JJ., Prota, AE., Kampmann, M., Akhmanova, A., Steinmetz, MO., Tanenbaum, ME., Weissman, JS. Combined CRISPRi/a-Based Chemical Genetic Screens Reveal that Rigosertib Is a Microtuble-Destabilization Agent. *Mole Cell*. **2017**. *68* (1), 210-223.

39. Deltcheva, E., Chylinski, K., Sharma, CM., Gonzales, K., Chao, Y., Pirzada, ZA., Eckert, MR., Vogel, J., Charpentier, E. CRISPR RNA maturation by *trans*-encoded small RNA and host factor RNase III. *Nature*. **2011**. *471*, 602-607.

40. Chylinski, K., Le Rhun, A., Charpentier, E. The tracrRNA and Cas9 families of type II CRISPR-Cas immunity systems. *RNA Biol*. **2013**. *10* (5), 726-737.

41. Zetsche, B., Gootenberg, JS., Abudayyeh, OO., Slaymaker, IM., Makarova, KS., Essletzbichler, P., Volz, SE., Joung, J., van der Oost, J., Regev, A., Koonin, EV., Zhang, F. Cpf1 Is a Single RNA-Guided Endonuclease of a Class 2 CRISPR-Cas System. *Cell*. **2015**. *163*, 759-771.

42. Fonfara, I., Richter, H., Bratovic, M., Le Rhun, A., Charpentier, E. The CRISPR-associated DNA-cleaving enzyme Cpf1 also processes precursor CRISPR RNA. *Nature*. **2016**. *532*, 517-521.

43. Zetsche, B., Heidenreich, M., Mohanraju, P., Fedorova, I., Kneppers, J., DeGennaro, EM., Winblad, N., Choudhury, SR., Abudayyeh, OO., Gootenberg, JS., Wu, WY., Scott, DA., Severinov, K., van der Oost, J., Zhang, F. Multiplex gene editing by CRISPR-Cpf1 using a single crRNA array. *Nat Biotechnol*. **2017**. *35*, 31-34.

44. Chow, RD., Kim, HR., Chen, S. Programmable sequential mutagenesis by inducible Cfp1 crRNA array inversion. *Nat Comm*. **2018**. *9*, 1903.

45. Chow, RD., Wang, G., Ye, L., Codina, A., Kim, HR., Shen, L., Dong, MB., Errami, Y., Chen, S. In vivo profiling of metastatic double knockouts through CRISPR-Cpf1 screens. *Nat Methods*. **2019**. *16*, 405-408.

46. Avery, S. Molecular targets of oxidative stress. *Biochem. J.* **2011.** *434*, 201-210.

47. Lesgards, J.; Gauthier, C.; Iovanna, J.; Vidal, N.; Dolla, A.; Stocker, P. Effect of reactive oxygen and carbonyl species on crucial cellular antioxidant enzymes. *Chem. Biol. Interact.* **2011.** *190*, 28-34.

48. Benzie, I. Evolution of antioxidant defense mechanisms. *Eur. J. Nutr.* **2000**. *39*, 53-61.

49. McCord, J. The Evolution of Free Radicals and Oxidative Stress. *Am. J. Med.* **2000**. *108*, 652-659.

50. Droge, W. Free Radicals in the Physiological Control of Cell Function. *Physiol Rev*. **2002**. *82*, 47-95.

51. Limon-Pacheco, J., Gonsebatt, ME. The role of antioxidants and antioxidant-related enzymes in protective resonses to environmentally induced oxidative stress. *Mutat Res*. **2009**. *674*, 137-147.

52. Finkle, T. Oxidant signals and oxidative stress. *Curr Opin Cell Biol*. **2003**. *15*, 247-254.

53. Leonard, SE., Carroll, KS. Chemical 'omics' approaches for understanding protein cysteine oxidation in biology. *Curr Opin Chem Biol*. **2011**. *15* (1), 88-102.

54. Klomsiri,C., Karplus, PA., Poole, LB. Cysteine-Based Redox Switches in Enzymes. *Antioxid Redox Sign*. **2011**. *14* (6), 1065-1077.

55. Reddie, KG., Carroll, KS. Expanding the functional diversity of proteins through cysteine oxidation. *Curr Opin Chem Biol*. **2008**. *12* (6), 746-754.

56. Paulsen, CE., Carroll, KS. Orchestrating Redox Signaling Networks through Regulatory Cysteine Switches. *ACS Chem Biol*. **2010**. *5*(1), 47-62.

57. Leonard, SE., Reddie, KG., Carroll, KS. Mining the Thiol Proteome for Sulfenic Acid Modifications Reveals New Targets for Oxidation in Cells. *ACS Chem Biol*. **2009**. *4* (9), 783-799.

58. Vandenbroucke, K., Robbens, S., Vandepoele, K., Inze, D., Van de Peer, Y., Breusegem, F. Hydrogen Peroxide-Induced Gene Expression across Kingdoms: A Comparative Analysis. *Mol. Biol. Evol.* **2008**. *25*, 507-516.

59. Gasch, A., Spellman, P., Kao, C., Carmel-Harel, O., Eisen, M., Storz, G., Botstein, D., Brown, P. Genomic Expression Programs in the Response of Yeast Cells to Environmental Changes. *Mol. Biol. Cell*. **2000**. *11*, 4241-4257.

60. Wei, L., Han, X., Xiao, T., Cong, L., Love, MI., Zhang, F., Irizarry, RA., Liu, JS., Brown, M., Liu, XS. MAGeCK enables robust identification of essential genes from genome-scale CRISPR/Cas9 knockout screens. *Genome Biol*. **2014**. *15* (554).

61. Eden, E., Navon, R., Steinfeld, I., Lipson, D., Yakhini, A. Gorilla: A Tool For Discovery And Visualization of Enriched GO Terms in Ranked Gene Lists. *BMC Bioinformatics*. **2009**. *10* (48).

62. Tom, CTMB, Crellin, JE., Motiwala, HF., Stone, MB., Davda, D., Walker,W., Kuo, YH., Hernandez, JL., Labby, KJ., Gomez-Rodriguez, L., Jenkins, PM., Veatch, SL., Martin, BR. Chemoseletive ratiometric imaging of protein S-sulfenylation. *Chem. Comm.* **2017**. *53*, 7385-7388.

63. Richard, E., Jorge-Finnigan, A., Garcia-Villoria, J., Merinero, B., Desviat, LR., Gort, L., Briones, P., Leal, F., Perez-Cerda, C., Ribes, A., Ugarte, M., Perez, B. Genetic and Cellular Studies of Oxidative Stress in Methylmalonic Aciduria (MMA) Cobalamin Deficiency Type C (*cblC*) With Homocystinuria (MMACHC). *Hum Mutat*. **2009**. *30* (11), 1558-1566.

64. Gong, J., Huang, M., Wang, F., Ma, X., Liu, H., Tu, Y., Xing, L., Zhu, X., Zheng, H., Fang, J., Li, X., Wang, Q., Wang, J., Sun, Z., Wang, X., Wang, Y., Guo, C., Tang, TS. RBM45 competes with HDAC1 for binding to FUS in response to DNA damage. *Nucleic Acids Res*. **2017**. *45* (22), 12862-12876.

65. Tanaka, E., Fukuda, H., Nakashima, K., Tsuchiya, N., Seimiya, H., Nakagama, H. HnRNP A3 binds to and protects mammalian telomeric repeats in vitro. *Biochem Biophys Res Commun.* **2007**. *358* (2), 608-614.

66. Gilpin, KM., Chang, L., Monteiro, MJ. ALS-linked mutations in ubiquilin-2 or hnRNPA1 reduce interaction between ubiquilin-2 and hnRNPA1. *Hum Mol Genet*. **2015**. *24* (9), 2565-2577.

67. Fifita, JA., Zhang, KY., Galper, J., Williams KL., McCann EP., Hogan, AL., Saunders, N., Bauer, D., Tarr, IS., Pamphlett, R., Nicholson, GA., Rowe, D., Yang, S., Blair, IP. Genetic and Pathological Assesment of hnRNPA1, hnRNPA2/B1, and hnRNPA3 in Familial and Sporadic Amyotrophic Lateral Sclerosis. *Neurodegener. Dis.* **2017**. *17*, 304-312.

68. Goswami, AV., Samaddar, M., Sinha, D., Purushotham, J., D'Silva, P. Enhanced J-protein interaction and compromised protein stability of mtHsp70 variants lead to mitochondrial dysfunction in Parkinson's disease. *Hum. Mol. Genet.* **2012**. *21* (15), 3317-3332.

69. Dubreuil, MM., Morgens, DW., Okumoto, K., Honsho, M., Contrepois, K., Lee-McMullen, B., Traber, GM., Sood, RS., Dixon, SJ., Snyder, MP., Fujiki, Y., Bassik, MC. Systematic Identification of Regulators of Oxidative Stress Reveals Non-canonical Roles for Peroxisomal Import and the Pentose Phosphate Pathway. *Cell Rep*. **2020**. *30*, 1417-1433.

70. Morgens, DW., Deans, RM., Li, A., Bassik, MC. Systematic comparison of CRISPR/Cas9 and RNAi screens for essential genes. *Nat*. *Biotechnol*. **2016**. *34*, 634-636.

71. Le Sage, C., Lawo, S., Panicker, P., Scales, TME., Rahman, SA., Little, AS., McCarthy, NJ., Moore, JD., Cross, BCS. Dual direction CRISPR transcriptional regulation screening uncovers gene networks driving drug resistance. *Sci. Rep*. **2017**. *7*(17693).

72. Hang, H.C. Exploring Protein Lipidation with Chemical Biology. *Chem Rev*. **2011**. *111*, 6341-6358.

73. Levental, I., Grzybek, M., Simons, K. Greasing Their Way: Lipid Modifications Determine Protein Association with Membrane Rafts. *Biochemistry*. **2010**. *49*, 6305-6316.

74. Lin, DTS., Conibear, E. ABHD17 proteins are novel protein depalmitoylases that regulate N-Ras palmitate turnover and subcellular localization. *eLife*. **2015**. *4*.

75. Won, SJ., Kit, MCS., Martin, BR. Protein Depalmitoylases. *Crit Rev Biochem Mole*. **2018**. *53* (1), 83-98.

76. De, I., Sadhukhan, S. Emerging Roles of DHHC-mediated Protein S-palmitoylation in Physiological and Pathophysiological Context. *Europen J. Cell Biol.* **2018**. *97*, 319-338.

77. Mitchell, DA., Vasudevan, A., Linder, ME., Deschenes, RJ. Protein palmitoylation by a family of DHHC protein *S*-acyltransferases. *J Lipid Res*. **2006**. *47*, 1118-1127.

78. Fukata, M., Fukata, Y., Adesnik, H., Nicoll, RA., Bredt, DS. Identification of PSD-95 Palmitoylating Enzymes. *Neuron*. **2004**. *44* (6), 987-996.

79. Politis, EG., Roth, AF., Davis, NG. Transmembrane Topology of the Protein Palmitoyl Transferase Akr1. *J Biol Chem*. **2005**. *280* (11), 10156-10163.

80. Lai, J., Linder, ME. Oligomerization of DHHC Protein *S*-Acyltransferases. *J Biol Chem*. **2013**. *288* (31), 22862-22870.

81. Jennings, BC., Linder, ME. DHHC Protein *S*-Acyltransferases Use Similar Ping-Pong Kinetic Mechanisms but Display Different Acyl-CoA Specificities. *J Biol Chem*. **2011**. *287* (10), 7236-7245.

82. Mitchell, DA., Mitchell, G., Ling, Y., Budde, C., Deschenes, RJ. Mutational Analysis of *Saccharomyces cerevisiae* Erf2 Reveals a Two-step Reaction Mechanism for Protein Palmitoylation by DHHC Enzymes. *J Biol Chem*. **2010**. *285* (49), 38104-38114.

83. Lobo, S., Greentrees, WK., Linder, ME., Deschenes, RJ. Identification of a Ras Palmitoyltransferase in *Saccharomyces cerevisiae*. *J Biol Chem*. **2002**. *277* (43), 41268-41273.

84. Ohno, Y., Kihara, A., Sano, T., Igarashi, Y. Intracellular localization and tissue-specific distribution of human and yeast DHHC cysteine-rich domain-containing proteins. *Biochimica et Biophysica Acta*. **2006**. *1761*, 474-483.

85. Lemonidis, K., Werno, MW., Greaves, J., Diez-Ardanuy, C., Sanchez-Perez, MC., Salaun, C., Thompson, DM., Chamberlain, LH. The zDHHC family of S-acyltransferases. *Biochem Soc Trans*. **2015**. *43* (2), 217-221.

86. Oku, S., Takahashi, N., Fukata, Y., Fukata, M. *In Silico* Screening for Palmitoyl Substrates Reveals a Role for DHHC1/3/10 (zDHHC1/3/11)-mediated Neurochondrin Palmitoylation in Its Targeting to Rab5-positive Endosomes. *J Biol Chem*. **2013**. *288* (27), 19816-19829.

87. Zhou ,Q., Lin, H., Wang, S., Ran, Y., Liu, Y., Ye, W., Xiong, X., Zhong, B., Shu, HB., Wang, YY. The ER-Associated Protein ZDHHC1 Is a Positive Regulator of DNA Virus-Triggered, MITA/STING-Dependent Innate Immune Signaling. *Cell Host & Microbe*. **2014**. *16*, 450-461.

88. Li, B., Cong, F., Tan, CP., Wang, SX., Goff, SP. Aph2, a Protein with a *zf*-DHHC Motif, Interacts with c-Abl and Has Pro-apoptotic Activity. *J Biol Chem.* **2002**. *277*, 28870-28876.

89. Zeidman, R., Buckland, G., Cevecauer, M., Eissmann, P., Davis, DM., Magee, AI. DHHC2 is a protein S-acyltransferase for Lck. *Mol Membr Biol*. **2011**. *28*, 473-486.

90. Greaves, J., Gorleku, OA., Salaun, C., Chamberlain, LH. Palmitoylation of the SNAP25 protein family: specificity and regulation by DHHC palmitoyl transferases. *J Biol Chem*. **2010**. *285* (32), 24629-24638.

91. Jia, L., Linder, ME., Blumeer, KJ. $G_{i/o}$ Signaling and the Palmitoyltransferase DHHC2 Regulate Palmitate Cycling and Shuttling of RGS7 Family-binding Protein. *J Biol Chem*. **2011**. *286* (15), 13695-13703.

92. Greaves, J., Carmichael, JA., Chamberlain, LH. The palmitoyl transferase DHHC2 targets a dynamic membrane cycling pathway: regulation by a C-terminal domain. *Mole Biol Cell*. **2011**. *22* (11), 1887-1895.

93. Gorleku, OA., Barns, AM., Prescott, GR., Greaves, J., Chamberlain, LH. Endoplasmic reticulum localization of DHHC palmitoyltransferases mediated by lysine-based sorting signals. *J Biol Chem*. **2011**. *286* (45), 39573-39584.

94. Wang, J., Hao, JW., Guo, H., Sun, HH., Lai, XY., Liu, LY., Zhu, M., Wang, HY., Li, YF., Yu, LY. Xie, C., Wang, HR., Mo, W., Zhou, HM., Chen, S., Liang, G., Zhou, TJ. DHHC4 and DHHC5 Facilitate Fatty Acid Uptake by Palmitoylating and Targeting CD36 to the Plasma Membrane. *Cell Rep*. **2019**. *26*, 209-221.

95. Ebersole, B., Petko, J., Woll, M., Murakami, S., Sokolina, K., Wong, V., Stagljar, I., Luscher, B., Levenson, R. Effect of C-Terminal S-Palmitoylation of D2 Dopamine Receptor Trafficking and Stability. *PLoS One*. **2015**. *10* (11).

96. Goytain, A., Hines, RM., Quamme, GA. Huntingtin-interacting proteins, HIP14 and HIP14L, mediate dual functions, palmitoyl acyltransferase and Mg2+ transport. *J Biol Chem*. **2008**. *283* (48), 33365-33374.

97. Singaraja, RR., Hadano, S., Metzler, M., Givan, S., Wellington, CL. Warby, S., Yanai, A., Gutekunst, CA., Leavitt, BR., Yi, H., Fichter, K., Gan, L., McCutcheon, K., Chopra, V., Michel, J., Hersch, SM., Iekda, JE., Hayden, MR. HIP14, a novel ankyrin domain-containing protein, links huntingtin to intracellular trafficking and endocytosis. *Hum Mole Genet*. **2002**. *11* (23), 2815-2828.

98. Huang, K., Sanders, S., Singaraja, R., Orban, P., Cijsouw, T., Arstikaitis, P., Yanai, A., Hayden, MR., El-Husseini, A. Neuronal palmitoyl acyl transferases exhibit distinct substrate specificity. *FASEB J*. **2009**. *23* (8), 2605-2615.

99. Adachi, N., Hess, DT., McLaughlin, P., Stamler, JS. S-Palmitoylation of a Novel Site in the β2-Adrenergic Receptor Associated with a Novel Intracellular Itinerary. *J Biol Chem*. **2016**. *291* (38), 20232-20246.

100. Zhang, F., Di, Y., Li, J., Shi, Y., Zhang, L., Wang, C., He, X., Liu, Y., Wan, D., Huo, K., Gu, J. Molecular cloning and characterization of human Aph2 gene, involved in AP-1 regulation by interaction with JAB1. *Biochimica et Biophysica Acta*. **2006**. *1759* (11-12).

101. Baumgart, F., Corral-Escariz, M., Perez-Gil, J., Rodriguez-Crespo, I. Palmitoylation of R-Ras by human DHHC19, a palmitoyl transferase with a CaaX box. *Biochimica et Biophysica Acta*. **2010**. *1798* (3), 592-604.

102. McMichael, TM., Zhang, L., Chemudupati, M., Hack, JC., Kenney, AD., Hang, HC., Yount, JS. The palmitoyltransferase ZDHHC20 enhances interferon-induced transmembrane protein 3 (IFITM3) palmitoylation and antiviral activity. *J Biol Chem*. **2017**. *292* (52), 21517-21526.

103. Pedram, A., Razandi, M., Deschenes, RJ., Levin, ER. DHHC-7 and -21 are palmitoyltransferases for sex steroid receptors. *Mole Biol Cell*. **2012**. *23* (1), 188-199.

104. Tian, L., McClafferty, H., Knaus, HG., Ruth, P., Shipston, MJ. Distinct acyl protein transferases and thioesterases control surface expression of calcium-activated potassium channels. *J Biol Chem*. **2012**. *287* (18), 14718-14725.

105. Jennings, BC., Nadolski, MJ., Ling, Y., Baker, MB., Harrison, ML., Deschenes, RJ., Linder, ME. 2-Bromopalmitate and 2-(2-hydroxy-5-nitro-benzylidene)-benzo[*b*]thiophen-3-one inhibit DHHC-mediated palmitoylation in vitro. *J Lipid Res*. **2009**. *50*, 233-242.

106. Ohno, Y., Kashio, A., Ogata, R., Ishitomi, A., Yamazaki, Y., Kihara, A. Analysis of substrate specificity of human DHHC protein acyltransferases using a yeast expression system. *Mole Biol Cell*. **2012**. *23* (23), 4473-4661.

107. Linder, ME., Jennings, BC. Mechanism and function of DHHC S-acyltransferases. *Biochem Soc Trans*. **2013**. *41*(1), 29-34.

108. Roth, AF., Wan, J., Bailey, AO., Sun, B., Kuchar, JA., Green, WN., Phinney, BS., Yates, JR., Davis, NG. Global Analysis of Protein Palmitoylation in Yeast. *Cell*. **2006**. *125* (5), 1003-1013.

109. Gottlieb, CD., Linder, ME. Structure and function of DHHC protein *S*-acyltransferases. *Biochem. Soc. Trans.* **2017**. *45*(4), 923-928.

110. Colicelli, J. Human RAS Superfamily Proteins and Related GTPases. *Science STKE*. **2004**. *250*.

111. Wennerberg, K., Rossman, KL., Der, CJ. The Ras superfamily at a glance. *J Cell Sci*. **2005**. *118*, 843-846.

112. Yeste-Velasco, M., Linder, ME., Lu, YJ. Protein S-palmitoylation and cancer. *Biochimica et Biophysica Acta*. **2015**. *1856*, 107-120.

113. Bishop, AL., Hall, A. Rho GTPases and their effector proteins. *Biochem J*. **2000**. *348*, 241-255.

114. Bernards, A., Settleman, J., GAP control: regulating the regulators of small GTPases. *Trends in Cell Biol*. **2004**. *14* (7), 377-385.

115. Repasky, GA., Chenette, EJ., Der, CJ. Renewing the conspiracy theory debate: doest Raf function alone to mediate Ras oncogenesis?. *Trends in Cell Biol*. **2004**. *14* (11), 639-647.

116. Samatar, AA., Poulikakos, PI. Targeting RAS-ERK signaling in cancer: promises and challenges. *Nat. Rev. Drug Discovery*. **2014**. *13*, 928-942.

117. Chenette, EJ., Der, CJ. Lipid Modification of Ras Superfamily GTPases: Not Just Membrane Glue. *The Enzymes*. **2011**. *29*, 59-95.

118. Misaki, R., Morimatsu, M., Uemura, T., Waguri, S., Miyoshi, E., Taniguchi, N., Matsuda, M., Taguchi, T. Palmitoylated Ras proteins traffic through recycling endosomes to the plasma membrane during exocytosis. *J Cell Biol*. **2010**. *191* (1), 23-29.

119. Rocks, O., Peyker, A., Kahms, M., Verveer, PJ., Koerner, C., Lumbierres, M., Khulmann, J., Waldmann, H., Wittinghofer, A., Bastiaens, PIH. An Acylation Cycle Regulates Localization and Activity of Palmitoylated Ras Isoforms. *Science*. **2005**. *307* (5716), 1746-1752.

120. Swarthout, JT., Lobos, S., Farh, L., Croke, MR., Greentree, WK., Deschenes, RJ., Linder, ME. DHHC9 and GPC16 constitute a human protein fatty acyltransferase with specificity for H- and N-Ras. *J. Biol. Chem*. **2005**. *280* (35), 31141-31148.

121. Song, SP., Hennig, A., Schubert, K., Markwart, R., Schmidt, P., Prior, IA., Bohmer, FD., Rubio, I. Ras palmitoylation is necessary for N-Ras activation and signal propagation in growth factor signaling. *Biochem J*. **2013**. 454, 323-332.

122. Hancock, JF. Ras proteins: different signals from different locations. *Nature*. **2003**. *4* (5), 373-384.

123. Lin, DTS., Davis, NG., Conibear, E. Targeting the Ras palmitoylation/depalmitoylation cycle in cancer. *Biochem. Soc. Trans.* **2017**. *45*, 913-921.

124. Chow, RD., Guangchuan, W., Lupeng, Y., Codina, A., Kim, HR., Shen, L., Dong, MB., Errami, Y., Chen, S. In Vivo profiling of metastatic double knockouts through CRISPR-Cpf1 screens. *Nat Methods*. **2019**. *16*, 405-408.

125. Bae, S., Park, J., Kim, J.S. Cas-OFFinder: A fast and versatile algorithm that searches for potential off-target sites of Cas9 RNA-guided endonucleases. *Bioinformatics*. **2014**. *30*, 1473-1475.

126. Bartha, I., di lulio, J., Venter, JC., Telenti, A. Human gene essentiality. *Nat Rev Genet*. **2018**. *19*, 51-62.

127. Liu, J., Srinivasan, S., Li, CY., Ho, IL., Rose, J., Shaheen, MA., Wang, G., Yao, W., Deem, A., Bristow, C., Hart, T., Draetta, G. Pooled library screening with multiplexed Cpf1 library. *Nat Comm*. **2019**. *10*, 1-10.

128. Ong, SE., Blagoev, B., Kratchmarova, I., Kristensen, DB., Steen, H., Pandey, A., Mann, M. Stable Isotope Labeling by Amino Acids in Cell Culture, SILAC, as a Simple and Accurate Approach to Expression Proteomics. *Mol Cell Proteomics*. **2002**. *1*, 376-386.

129. Forrester, MT., Hess, DT., Thompson, JW., Hultman, R., Moseley, MA., Stamler, JS., Casey, PJ. Site-Specific analysis of protein *S*-acylation by resin-assisted capture. *Journal of Lipid Research*. **2011**. *52*, 393-398

130. Hsu, PD., Scott, DA., Weinstein, JA., Ran, FA., Konermann, S., Agarwala, V., Li Y., Fine, EJ., Wu, X., Shalem, O., Cradick, TJ., Marraffini, LA., Bao, G., Zhang, F. DNA targeting specificity of RNA-guided Cas9 nucleases. *Nat Biotech*. **2013**. *31* (9), 827-832.

131. Doench, JG., Fusi, N., Sullender, M., Hegde, M., Vaimber, EW., Donovan, KF., Smith, I., Tothova, Z., Wilen, C., Orchard, R., Virgin, HW., Listgarten, J., Root, DE. Optimized sgRNA design to maximize activity and minimize off-target effects of CRISPR-Cas9. *Nat Biotech*. **2016**. *34* (2), 184-191.

132. Doench, JG., Hartenian, E., Graham, DB., Tothova, Z., Hegde, M., Smith, I., Sullender, M., Ebert, BL., Xavier, RJ., Root, DE. Rational design of highly active sgRNAs for CRISPR-Cas9-mediated gene inactivation. *Nat Biotech*. **2014**. *32* (12), 1262-1267.

133. Cui, Y., Xu, J., Cheng, M., Liao, X., Peng, S. Review of CRISPR/Cas9 sgRNA Design Tools. *Interdiscip Sci*. **2018**. *10*, 455-465.

134. Hart, T., Tong, AHY., Chan, K., Leeuwen, JV., Seetharaman, A., Aregger, M., Chandrashekhar, M., Hustedt, N., Seth, S., Noonan, A., Habsid, A., Sizova, O., Nedyalkova, L., Climie, R., Tworzyanski, L., Lawson, K., Sartori, MA., Alibeh, S., Tieu, D., Masud, S., Mero, P., Weiss, A., Brown, KR., Usaj, M., Billmann, M., Rahman, M., Costanzo, M., Myers, CL., Andrews, BJ., Boone, C., Durocher, D., Moffat, J. Evaluation and Design of Genome-Wide CRISPR/SpCas9 Knockout Screens. *G3*. **2017**. *7*, 2719-2727.

135. Brisoy, K., Wang, T., Chen, WW., Freinkman, E., Abu-Remaileh, M., Sabatini, DM. An Essential Role of the Mitochondrial Electron Transport Chain in Cell Proliferation Is to Enable Aspartate Synthesis. *Cell*. **2015**. *162* (3), 540-551.

136. Tarumoto, Y., Lu, B., Somerville, TDD., Huang, YH., Milazzo, JP., Wu, XS., Klingbeil, O., El Demerdash, O., Shi, J., Vakoc, CR. LKB1, Salt-Inducible Kinases, and MEF2C are Linked Dependencies in Acutre Myeloid Leukemia. *Mole Cell*. **2018**. *69* (6), 1017-1027.

137. Huang, YH., Klingbeil, O., He, XY., Wu, XS., Arun, G., Lu, B., Somerville, TDD., Milazzo, JP., Wilkinson, JE., Demerdash, OE., Spector, DL., Egeblad, M., Shi, J., Vakoc, CR. POU2F3 is a master regulator of a tuft cell-like variant of small cell lung cancer. *Genes Dev*. **2018**. *32* (13-14), 915-928.

APPENDIX A

Cloning, Expression and Validation of pCLHCX-AsCpf1

To circumvent potential problems with expression of the Cpf1 nuclease in MEF cell models already under blastocidin (HRas isoforms) and puromycin selection (gRNA libraries), AsCpf1 was subcloned into a pCLHCX vector containing a hygromycin selection cassette. Primers specific to AsCpf1 containing Hindiii and NotI restriction sites were used to amplify the gene from pY108 (Addgene #84739).

The following primers and conditions were used:

pCL-AsCpf1_F :

TTTAAGCTTGCCACCATGGCCCCAAAGAAGAAG

pCL-AsCpf1_R:

TTTGCGGCCGCCTTTTTCTTTTTTGCCTGGC

Plasmid DNA :	100ng
$MgCl_2$ (50mM) :	1µL
dNTP Mix (10mM):	1µL
pCL-AsCpf1_F :(10µM)	2.5µL
pCL-AsCpf1_R (10µM)	2.5µL
Phusion Buffer (5x):	10µL
Phusion High-Fidelity Polymerase (NEB):	1µL
diH_2O:	to 50µL

Thermocycler conditions:

1. 98°C 1min.
2. 98°C 30s.
3. 55°C 30s.
4. 74°C 20s. ➔ repeat steps 2-4 for 34 cycles
5. 74°C 10min.
6. 4°C ∞

The PCR product was confirmed by running the reaction on a 1% agarose gel (≈4400 bp amplicon) and extracted with a QIAquick Gel Extraction Kit.

The pCLHCX parent vector and AsCpf1 insert were then restriction digested to generate compatible ends for ligation:

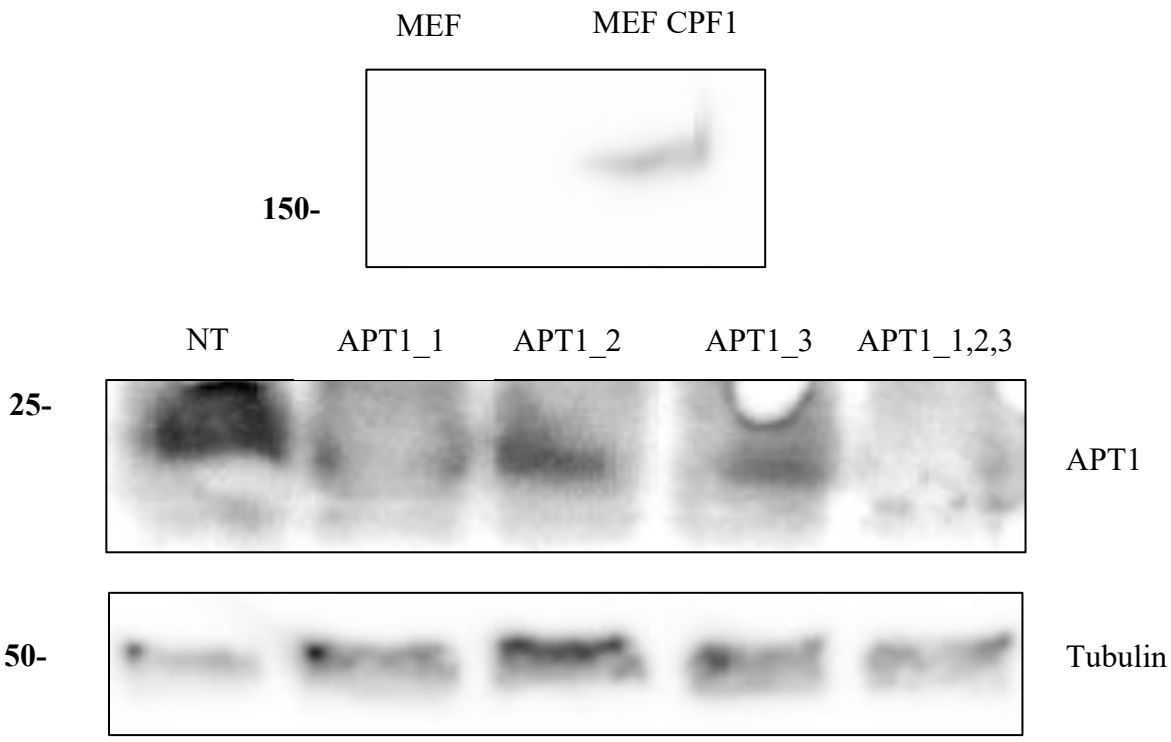
Plasmid DNA :	1µg
Gene Insert :	500ng
Hindiii	1µL
NotI	1µL
NEB Cutsmart Buffer (10x)	10µL
diH_2O:	to 50µL

The digest was allowed to run for 90 minutes at 37°C and collected using a Zymo DNA concentrator column and following the manufacturer's instructions. The purified Cpf1 fragment was then ligated into the linearized pCLHCX vector:

Plasmid DNA :	50ng
Gene Insert :	25ng
T4 DNA Ligase Buffer (5x)	10µL
T4 DNA Ligase(Thermo)	1µL
diH_2O:	to 50µL

The ligation was allowed to run overnight and the assembled vector collected using a Zymo DNA concentrator column. The ligated vectors were then transformed into Stbl3 electrocompetent cells and colonies selected for validation. Once cloning of AsCp1was

confirmed, it was packaged into retrovirus and used to transduce HRas-dependent MEF cell lines to confirm expression and activity. Anti-His 6x western blot of cell lysates reveled expression a single protein at the expected molecular weight of ≈165kDa. Single gRNAs targeted to cellular thioesterase APT1 were cloned into pU6-sgRNA EF1Alpha-puro-T2A-BFP (Addgene #60955) and stably transduced into MEF-HRas cultures expressing AsCpf1 along with a control lacking a targeting sequence. Western blot of APT1 was then used to assess the efficacy of AsCpf1.



APPENDIX B

PCR Primers for Illumina Library Preparation

<u>Toronto Knockout Library (TKOv1)</u>

Nested PCR Step One primer sequences

Outer Primer Forward: AGGGCCTATTTCCCATGATTCCTT

Outer Primer Reverse: TCAAAAAAGCACCGACTCGG

Nested PCR Step Two primer sequences

Illumina True seq adapters with i5 barcodes

Forward 1:
AATGATACGGCGACCACCGAGATCTACAC**TATAGCCT**ACACTCTTTCCCTACACGA
CGCTCT TCCGATCT**TGTGGAAGGACGAGGTACCG**

Forward 2:
AATGATACGGCGACCACCGAGATCTACAC**ATAGAGGC**ACACTCTTTCCCTACACGA
CGCTCT TCCGATCT**TGTGGAAGGACGAGGTACCG**

Forward 3:
AATGATACGGCGACCACCGAGATCTACAC**CCTATCCT**ACACTCTTTCCCTACACGA
CGCTCT TCCGATCT**TGTGGAAGGACGAGGTACCG**

Forward 4:
AATGATACGGCGACCACCGAGATCTACAC**GGCTCTGA**ACACTCTTTCCCTACACGA
CGCTCT TCCGATCT**TGTGGAAGGACGAGGTACCG**

Illumina True seq adapters with i7 barcodes

Reverse 1:
CAAGCAGAAGACGGCATACGAGAT**CGAGTAAT**GTGACTGGAGTTCAGACGTGTGCT
CTTCC GATC**TATTTTAACTTGCTATTTCTAGCTCTAAAAC**

Reverse 2:
CAAGCAGAAGACGGCATACGAGAT**TCTCCGGA**GTGACTGGAGTTCAGACGTGTGCT
CTTCC GATC**TATTTTAACTTGCTATTTCTAGCTCTAAAAC**

Reverse 3:
CAAGCAGAAGACGGCATACGAGAT**AATGAGCG**GTGACTGGAGTTCAGACGTGTGC
TCTTCC GATC**TATTTTAACTTGCTATTTCTAGCTCTAAAAC**

Reverse 4:
CAAGCAGAAGACGGCATACGAGAT**GGAATCTC**GTGACTGGAGTTCAGACGTGTGCT
CTTCC GATC**TATTTTAACTTGCTATTTCTAGCTCTAAAAC**

Reverse 5:
CAAGCAGAAGACGGCATACGAGAT**TTCTGAAT**GTGACTGGAGTTCAGACGTGTGCT
CTTCCG ATC**TATTTTAACTTGCTATTTCTAGCTCTAAAAC**

Reverse 6:
CAAGCAGAAGACGGCATACGAGAT**ACGAATTC**GTGACTGGAGTTCAGACGTGTGCT
CTTCC GATC**TATTTTAACTTGCTATTTCTAGCTCTAAAAC**

Red: barcode

Bold: constant vector region

Index keys for sequencing sample submission sheet:

		I5 Index	I5 Index for Submission			I7 Index	I7 Index for Submission
F1	D501	TATAGCCT	TATAGCCT	R1	D701	CGAGTAAT	ATTACTCG
F2	D502	ATAGAGGC	ATAGAGGC	R2	D702	TCTCCGGA	TCCGGAGA
F3	D503	CCTATCCT	CCTATCCT	R3	D703	AATGAGCG	CGCTCATT
F4	D504	GGCTCTGA	GGCTCTGA	R4	D704	GGAATCTC	GAGATTCC
F5	D505	AGGCGAAG	AGGCGAAG	R5	D705	TTCTGAAT	ATTCAGAA
F6	D506	TAATCTTA	TAATCTTA	R6	D706	ACGAATTC	GAATTCGT

mDHHC/hDHHC Dual Knockout Libraries

Nested PCR Step One primer sequences

Outer Primer Forward (DHHC_vrd1_F):

AATGGACTATCATATGCTTACCGTAACTTGAAAGTATTTCG

Outer Primer Reverse (DHHC_vrd1_R):

CTTTAGTTTGTATGTCTGTTGCTATTATGTCTACTATTCTTTCCC

Nested PCR Step Two primer sequences

Illumina Forward Adapter Sequences

Forward 1:

AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATC
TTAAGTAGAG**TCTTGTGGAAAGGACGAAACACCG**

Forward 2:

AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATC
TATACACGATC**TCTTGTGGAAAGGACGAAACACCG**

Forward 3:

AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATC
TGATCGCGCGGT**TCTTGTGGAAAGGACGAAACACCG**

Forward 4:

AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATC
TCGATCATGATCG**TCTTGTGGAAAGGACGAAACACCG**

Forward 5:

AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATC
TTCGATCGTTACCA**TCTTGTGGAAAGGACGAAACACCG**

Forward 6:

AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATC
TATCGATTCCTTGGT**TCTTGTGGAAAGGACGAAACACCG**

Forward 7:

AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATC
TGATCGATAACGCATT**TCTTGTGGAAAGGACGAAACACCG**

Forward 8:

AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATC
TCGATCGATACAGGTAT**TCTTGTGGAAAGGACGAAACACCG**

Forward 9:

AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATC
TACGATCGATAGGTAAGG**TCTTGTGGAAAGGACGAAACACCG**

Forward 10:

AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATC
TTAACAATGG**TCTTGTGGAAAGGACGAAACACCG**

Forward 11:

AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATC
TATACTGTATC **TCTTGTGGAAAGGACGAAACACCG**

Forward 12:

AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATC
TGATAGGTCGCA **TCTTGTGGAAAGGACGAAACACCG**

Illumina Reverse Adapter Sequences

Reverse 1:

CAAGCAGAAGACGGCATACGAGATAAGTAGAGGTGACTGGAGTTCAGACGTGTGCT
CTTCCGATCTT**TCTACTATTCTTTCCCCTGCACTGT**

Reverse 2:

CAAGCAGAAGACGGCATACGAGATACACGATCGTGACTGGAGTTCAGACGTGTGCT
CTTCCGATCTAT**TCTACTATTCTTTCCCCTGCACTGT**

Reverse 3:

CAAGCAGAAGACGGCATACGAGATCGCGCGGTGTGACTGGAGTTCAGACGTGTGCT
CTTCCGATCTGAT**TCTACTATTCTTTCCCCTGCACTGT**

Reverse 4:

CAAGCAGAAGACGGCATACGAGATCATGATCGGTGACTGGAGTTCAGACGTGTGCT
CTTCCGATCTCGAT**TCTACTATTCTTTCCCCTGCACTGT**

Reverse 5:

CAAGCAGAAGACGGCATACGAGATCGTTACCAGTGACTGGAGTTCAGACGTGTGCT
CTTCCGATCTCGATC**TCTACTATTCTTTCCCCTGCACTGT**

Reverse 6:

CAAGCAGAAGACGGCATACGAGATTCCTTGGTGTGACTGGAGTTCAGACGTGTGCT
CTTCCGATCTT**TCTACTATTCTTTCCCCTGCACTGT**

Reverse 7:

CAAGCAGAAGACGGCATACGAGATAACGCATTGTGACTGGAGTTCAGACGTGTGCT
CTTCCGATCTAT**TCTACTATTCTTTCCCCTGCACTGT**

Reverse 8:

CAAGCAGAAGACGGCATACGAGATACAGGTATGTGACTGGAGTTCAGACGTGTGCT
CTTCCGATCTGAT**TCTACTATTCTTTCCCCTGCACTGT**

Reverse 9:

CAAGCAGAAGACGGCATACGAGATAGGTAAGGGTGACTGGAGTTCAGACGTGTGCT
CTTCCGATCTCGAT**TCTACTATTCTTTCCCCTGCACTGT**

Reverse 10:

CAAGCAGAAGACGGCATACGAGATAACAATGGGTGACTGGAGTTCAGACGTGTGCT
CTTCCGATCTCGATC**TCTACTATTCTTTCCCCTGCACTGT**

Reverse 11:

CAAGCAGAAGACGGCATACGAGATACTGTATCGTGACTGGAGTTCAGACGTGTGCT
CTTCCGATCTT**TCTACTATTCTTTCCCCTGCACTGT**

Reverse 12:

CAAGCAGAAGACGGCATACGAGATAGGTCGCAGTGACTGGAGTTCAGACGTGTGCT
CTTCCGATCTAT**TCTACTATTCTTTCCCCTGCACTGT**

Red: barcode

Bold: constant vector region

Nested PCR hDHHC Step One primer sequences

Outer Primer Forward (DHHC_vrd1_F):

AATGGACTATCATATGCTTACCGTAACTTGAAAGTATTTCG

Outer Primer Reverse (hDHHC_R1):

GTAATTCTTTAGTTTGTATGTCTGTTGCTATTATGTCTACTATTC

Illumina hDHHC Forward Adapter Sequences

Forward 1:

AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATC
TTAAGTAGAG**TCTTGTGGAAAGGACGAAACACCG**

Forward 2:

AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATC
TATACACGATC**TCTTGTGGAAAGGACGAAACACCG**

Forward 3:

AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATC
TGATCGCGCGGT**TCTTGTGGAAAGGACGAAACACCG**

Forward 4:

AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATC
TCGATCATGATCG**TCTTGTGGAAAGGACGAAACACCG**

Forward 5:

AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATC
TTCGATCGTTACCA**TCTTGTGGAAAGGACGAAACACCG**

Forward 6:

AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATC
TATCGATTCCTTGGT**TCTTGTGGAAAGGACGAAACACCG**

Forward 7:

AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATC
TGATCGATAACGCATT**TCTTGTGGAAAGGACGAAACACCG**

Forward 8:

AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATC
TCGATCGATACAGGTAT**TCTTGTGGAAAGGACGAAACACCG**

Forward 9:

AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATC
TACGATCGATAGGTAAGG**TCTTGTGGAAAGGACGAAACACCG**

Forward 10:

AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATC
TTAACAATGG**TCTTGTGGAAAGGACGAAACACCG**

Forward 11:

AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATC
TATACTGTATC **TCTTGTGGAAAGGACGAAACACCG**

Forward 12:

AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATC
TGATAGGTCGCA **TCTTGTGGAAAGGACGAAACACCG**

Illumina hDHHC Reverse Adapter Sequences

Reverse 1:

CAAGCAGAAGACGGCATACGAGATAAGTAGAGGTGACTGGAGTTCAGACGTGTGCT
CTTCCGATCTT**TTATGTCTACTATTCTTTCCCCTGC**

Reverse 2:

CAAGCAGAAGACGGCATACGAGATACACGATCGTGACTGGAGTTCAGACGTGTGCT
CTTCCGATCTAT**TTATGTCTACTATTCTTTCCCCTGC**

Reverse 3:

CAAGCAGAAGACGGCATACGAGATCGCGCGGTGTGACTGGAGTTCAGACGTGTGCT
CTTCCGATCTGAT**TTATGTCTACTATTCTTTCCCCTGC**

APPENDIX C

zDHHC sgRNA Sequences

zDHHC	sgRNA #	Strand	Sequence	PAM
zDHHC1	**1**	+	GGGTCCTTGTTCCCCTCTTG	TTTG
	2	+	CTGGCCACCTTGTGGTGCAC	TTTG
	3	-	GATCGCGCACTCACGTCCAC	TTTG
	4	+	TACACAGTGTGGCATCTGCT	TTTC
zDHHC2	**1**	-	CCATGGACACTATGCACAGC	TTTC
	2	-	CCAGTATGACCAGACAAACA	TTTT
	3	-	GCTGCTCGCCTCAGAACTTC	TTTG
	4	-	TCACAGACGGAACAATGATG	TTTA
zDHHC3	**1**	-	CGCTCAATGTCTCGGAAGTG	TTTC
	2	+	TCCGAGATGGCTGTGGCATT	TTTA
	3	+	TCATGCTGGTTCCATCCCGA	TTTG
	4	-	GGCACTGCCCCGGGGTCCGT	TTTG
zDHHC4	**1**	-	GTGAAGATGCAGATCAGGAC	TTTT
	2	+	AAGGCCGTGGTCCTTGGAGG	TTTG
	3	-	GTGATGGTACCAGGATTGGC	TTTA
	4	+	AGGAAACCAGCCCGCTCCAA	TTTA
zDHHC5	**1**	-	AACCTCTTTCCAGACTCTGC	TTTG
	2	+	CGTGTCCAGGACTAAGCCTG	TTTA
	3	+	CCCGAGCCGAGGAAGATGAA	TTTC
	4	+	CCGCCCTCCTAGATGCTCCC	TTTA
zDHHC6	**1**	-	CAAACTTGATTACTGAGCAG	TTTT
	2	+	CCGCCACCTTGTTTGCCTTG	TTTG
	3	-	GCCTTTTCTTCAATCCATGA	TTTA
	4	+	CATGGTCGGGAGTTCCTGAA	TTTA
zDHHC7	**1**	-	GGGACAGCACCCGGGTCAGT	TTTG
	2	+	CAAGAGATGCATTCGAAAGA	TTTG
	3	+	ACCATGTACATAGCTCTGTC	TTTC
	4	-	ATCATTGCATATCGAGTGGA	TTTC
zDHHC8	**1**	-	AGGCGCGTCCCGGGGCTGCG	TTTG
	2	+	TCCTGGCCAACTTCAGTATG	TTTG
	3	+	ACCGTCCACCGCGCTGCTCA	TTTT
	4	+	ACCATCACTGCCCCTGGGTC	TTTG
zDHHC9	**1**	-	CGTGTCACCTTCTTTCTTAC	TTTC

	2	**-**	CCGGGCCATCATGACGCGGC	TTTG
	3	+	CCTTCGAGTGTCGCTACCTG	TTTG
	4	**-**	TCTGGAAATTCTTAATACGA	TTTA
zDHHC11	**1**	**-**	CTGGGCTTCTGGGAGGACCC	TTTC
	2	+	AGGCAATATCCTGGATCACC	TTTC
	3	**-**	GCTTTTTTGCTCGCGGTAAC	TTTG
	4	**-**	CAGTGGTGGTCAAAGCCAGA	TTTG
zDHHC12	**1**	+	GCCGCCGCTTAACTCTGGGA	TTTG
	2	+	TCTGGCTGTGTCACTCATGG	TTTA
	3	+	TGGCCTACCTGGCACTGCAG	TTTG
	4	**-**	GAGTCGGCCCTCGGTCAAAG	TTTC
zDHHC13	**1**	**-**	TCCCTTAGCATTTCGGATGT	TTTC
	2	+	CAAGGTTTTTGGTTGGCTAT	TTTC
	3	+	GCACATCATGTCTTATAAGG	TTTT
	4	+	ATCAGCACTGCTTTTGGACT	TTTG
zDHHC14	**1**	**-**	ATGGGCCCGCCGCCGCCGGG	TTTC
	2	**-**	CTCCGGGCCGCGATTTTCTT	TTTC
	3	+	GCAACGGGAGGATCATGATG	TTTT
	4	**-**	ACGGTCTGGCCATTGATGAC	TTTC
zDHHC15	**1**	+	AACTCTGCCTGGTGACTGTT	TTTG
	2	**-**	ATAGCGCTCCTTGTCTGTAT	TTTT
	3	**-**	AAACACACATGGCACAGACA	TTTA
	4	**-**	TAGTTGGAAAATCCAATGCA	TTTG
zDHHC16	**1**	+	GCTCCTGGGCTACCGGCGTC	TTTT
	2	+	CCTGCGCTCCCTGCTCTACA	TTTG
	3	+	GAGTGGTGTTTGTGGTGCTG	TTTG
	4	**-**	GCCAGGGGCAGTGATGATCC	TTTA
zDHHC17	**1**	**-**	GCTATACAGCGGTTGCACAC	TTTG
	2	+	TGCCGAAATTCCTTGTCAGC	TTTA
	3	**-**	GTCAAATACTATATTTCGAA	TTTC
	4	**-**	ATTTCCTCCGGGTGGAGAAG	TTTC
zDHHC18	**1**	+	ACGGACCCTGGCATCCTGCC	TTTC
	2	**-**	ACTGTCTGCCCGTTGATCAT	TTTC
	3	+	GGCCTCCCCGGACCTCGCAC	TTTC
	4	**-**	ATATCTTCGTTAGTTGTCAG	TTTG
zDHHC19	**1**	+	ATGGTTCCCCTCGAGCGTGT	TTTC
	2	+	TTCGGATTCCCTTGTAGGTG	TTTC
	3	**-**	GTGGAGCCTCGATGTAAGAT	TTTG
	4	+	CGTGGAGGACTTCGACCACC	TTTG
zDHHC20	**1**	+	TGGTCGTCTGGTCCTACTAC	TTTG
	2	+	CAGAACAGGAGAGAAAGGAA	TTTC
	3	**-**	ACAAGTAGAACTCTTTGGAA	TTTG
	4	**-**	TCACAATATCGGATAGCCTT	TTTT
zDHHC21	**1**	+	TTGTTGACCCGCATGGTTGG	TTTG
	2	+	CTCACTATGAAGAAGGACAT	TTTC

	3	+	GCCTGGTTGCCTTAGTGAGG	TTTT
	4	-	TCCAAGGACAGTGATGGTCC	TTTA
zDHHC22	**1**	+	CCTGGTGACCTTCGTACTGC	TTTC
	2	+	TTCACCGGCAACTGCATCGG	TTTC
	3	+	CTCCATGGTGGCTGGAGTGG	TTTA
	4	+	CTGTCGGCCTGGCCTGCGCC	TTTG
zDHHC23	**1**	-	TTTTTTTTGACTGGCTTCAT	TTTC
	2	-	GCAAGTCTCCGGCCGAAGTG	TTTC
	3	-	GACCGGGCATTCAATTTGGC	TTTT
	4	-	TCCAGACACAATGATGGTCC	TTTA
zDHHC24	**1**	-	GCACTGGTAACAGTAAGCCC	TTTG
	2	-	CCTGTGAGTAGCATGAGCCA	TTTG
	3	+	TGGTGGACACGTGTGTGGCA	TTTG
	4	+	CATGGGATGCTGCTGCTTCG	TTTC
zDHHC25	**1**	+	ACAATGTACATTGGACTTAC	TTTC
	2	+	TCCTCCTGGTGGCGATAATG	TTTC
	3	+	CCGTGGTGATGCTCTGCAGC	TTTG
	4	-	CGCACTTGGACCATCTGCCC	TTTG

Targeted Control sgRNAs

Ras Isoform	sgRNA #	Strand	Sequence	PAM
NRas	**1**	-	CTGTAAGAATCCTCTATGGT	TTTT
	2	+	TTGGACATACTGGATACAGC	TTTG
	3	-	CCACTAGCACCATAGGTACA	TTTC
	4	-	AATGAATGGAATCCCGTAAC	TTTC
HRas	**1**	+	GTGGACGAATACGACCCCAC	TTTT
	2	+	CCATCAACAACACCAAGTCT	TTTG
	3	-	ATCTGCTCCCTGTACTGGTG	TTTG
	4	+	GAGGACATCCACCAGTACAG	TTTT
KRas4a	**1**	+	GTGGACGAATATGATCCAAC	TTTT
	2	-	CTACTAGGACCATAGGTACA	TTTC
	3	-	AATCTGTATTGTCGGATCTC	TTTC
	4	+	TACATTGGTGAGGGAGATCC	TTTA
KRas4b	**1**	-	TGTTTTCGAATTTCTCGAAC	TTTA
	2	-	TACCATCTTTGCTCATCTTT	TTTT

Non-targeted Control sgRNAs

#	0	10	20	30	40	50
0		TATCGAA CGCTATC GAACGT	CGCGACG GCGGTTTT AACAC	TGGCGCG TAATCGCT ATTCG	AGCGTCG TACGATTT ATCGT	CGCCGAA CGGATAA GACTAT
01	TATAGCG TCGCGCG CTAATT	TCCGGTTA CGACCGG CAAAA	GTCGCGTT GCGACCG GCGCT	CAACCCG CCGTAGC GGGCCT	CGGAACG CGTGATA CTAGGG	CGACCGA GGTTACG CTGGGA
02	CGCGTAA CCGTTCC GGCTCC	CGTCGTA GAACGAT ACCCGA	ACCGCTA ATCGGCG TAAACT	TTTGCGTT GACGCGA TTACT	AGCGGGT CGTAGAC GCGCGG	ATCGCCG TACGGCT ATGGCA
03	CGCTAAG TTCGCGC GGGCCG	TTATGTCG CGCCGCG ACCCG	CGCGATC TATACCGT CTATA	AACGGAG CGCGATA CTACCA	TCTCGTAT CGCGATT GTTAA	
04	CGTTCGTC GACGAAG CAATC	ATTCCGG CGCGACC AGAAGA	TCGACGC GAGACGA CGTACA	CCGGACG TTATAAC GTTTCT	CGGTCGA CTACGGCT GTTAG	
05	GACCGTT GTATCGC GATCAA	GTATAAC TTCGCGC GGTCCG	CGATCGT ACGGGTG TCCTAA	ATACCCG TCGCGAA CGTGAG	TCGGGGC GCGTCAA CTGAAT	
06	GACCGTT GTATCGC GATCAA	GCGCGTC AACGTAA TGTGCC	GCCGACC GCGTGAT TGTAGG	ATTCGCG AACCCGG CCGGAA	GCTTACGC GTCGCGA ATGTA	
07	CCGTTCGT GCGATAG TGTTT	ACGCGCT ATTATCGT AGGTG	GCACGTT ACGCCCG GGTCAC	TTGATGG CGCGACG CGTTAT	ATTGCCGC TCGCGCA AAATC	
08	GTTGTCGT GCCGCGA ATTTA	ACTATCG ACGTTCGT CCCGG	TAAGCGA CCGACGC ATTCGC	GGTCAAA TCGCGGC GATAAC	TAACGCG CAACACG CCAATG	
09	CGGTACG TTAGACG AAAATT	TCGCGTTA TCGTGCC GTGCG	TAGCGTTC GACACGT GAAAC	CGACTCG CGCATTG AACTGG	CGAATCG CGCCTAC GGGAAT	

APPENDIX D

zDHHC sgRNA Sequences

zDHHC	sgRNA #	Strand	Sequence	PAM
zDHHC1	**1**	+	CTGTGATCGGCTTTGGGATC	TTTG
	2	+	GGATCCTTGTTCCCCTCCTG	TTTG
	3	+	CTGGCCACCTTGTGGTGCAC	TTTG
	4	+	GACCACCACTGCAAGTGGCT	TTTC
zDHHC2	**1**	-	CATGGACACTATGCACAGCT	TTTC
	2	+	GCAATGTTTGTCTGGTCATA	TTTT
	3	-	TAAGTTGGCATCTGTCACAG	TTTA
	4	+	TTGGAGAGAGAGCCAAGAGG	TTTG
zDHHC3	**1**	-	CGCTCAATGTTTCGGAAGTG	TTTC
	2	-	GGCACTGCCCCGGGGTCCGT	TTTG
	3	+	TCCGTGACGGCTGTGGCATC	TTTA
	4	-	GGGCACTTGTACACCACCTG	TTTG
zDHHC4	**1**	-	GAGCAGACGCAGATAAGAAC	TTTC
	2	+	CCATACGAGAAACCACACCT	TTTT
	3	+	TCACCCTGACTTGTGGAACC	TTTT
	4	+	AGGAAACCAGCTCGATCCAA	TTTA
zDHHC5	**1**	+	CGTGTCCAGGACTAAGCCTG	TTTA
	2	+	CCCTCGAGCTGAGGAGGATG	TTTT
	3	+	CGAGCTCCCCTTTACAAAAC	TTTC
	4	+	ACCGTCCCCCTCGATGTTCC	TTTT
zDHHC6	**1**	-	CAAACTTGATAACCGAACAG	TTTT
	2	-	CGGTTTCCACCCCAGAGGGA	TTTC
	3	+	TCCCTCTGGGGTGGAAACCG	TTTG
	4	-	TGCAGTGATGTGAACGTGGT	TTTC
zDHHC7	**1**	-	GAAGGCAGCAGCATGACGAA	TTTG
	2	-	GGTACTGCCCCAGGGTCGGT	TTTG
	3	-	ATACAGCAGCACTTGGGGCA	TTTA
	4	-	GCAAATACTGCAGTGGTGGG	TTTT
zDHHC8	**1**	-	AGGCGCGTCCCGGGGCTGCG	TTTG
	2	+	TCCTGGCCAACTTCAGCATG	TTTG
	3	+	ATGGACCCTGGTGTTTTCCC	TTTC
	4	+	CCCCCGAGCGGATGAGGATG	TTTT
zDHHC9	**1**	-	CCGGGCCATCATGACGCGGC	TTTG
	2	+	AGTGCCGCTACCTGGCTGTT	TTTG
	3	-	TATGAAAGCTGCTTCATCTG	TTTC

	4	**-**	TCTGGAAATTCTTGATACGA	TTTA
zDHHC11	**1**	**-**	CATTATTGAGTATGGCTTCT	TTTT
	2	+	CTCGGCCACCTTCGGGATCT	TTTC
	3	**-**	CTTGTTCACGGTGACCTTGC	TTTT
	4	+	TTGGCTTGGTGCACCTGGGC	TTTC
zDHHC11B	**1**	+	CTTGGCCACCTTCAGGATCT	TTTC
	2	**-**	TTGTTCACGGTGACCTTGCA	TTTC
	3	+	TCCAGCAAGGAGCTGGCGCC	TTTC
	4	**-**	TTGTTGCAGCCCAAGTGTAG	TTTC
zDHHC12	**1**	**-**	AGCTCCTCCTGAGGCTGGGG	TTTG
	2	+	TGGTCTACCTGGCGCTGCAG	TTTG
zDHHC13	**1**	+	GTCGATATGGCATCTGTGCA	TTTG
	2	+	CAAGGTTCTTGGTTGGGTAT	TTTC
	3	+	CCTGATTTAGCAGGAGCCCC	TTTC
	4	+	GTACATCATGTCTTATAAGG	TTTT
zDHHC14	**1**	**-**	ATGGGCCCGCCGCCGCCGGG	TTTC
	2	**-**	ACCGCCAGGTACGGACAGTC	TTTC
	3	**-**	GTTCTGGGAGGCGGGCGGTA	TTTG
	4	+	TGATAACTGCGTAGAACGGT	TTTG
zDHHC15	**1**	**-**	CTGCTGGGCTCAAAACAGTC	TTTT
	2	**-**	GCTGCTGTGGGAGTGTAAAG	TTTG
	3	**-**	GCCATATCAACAAGCATCTG	TTTG
	4	**-**	ACACACACATAGCACAGACA	TTTA
zDHHC16	**1**	+	GGGGCAGTGACACCGCTGTT	TTTG
	2	+	GAGTGGTGTTCGTGGTCCTG	TTTG
	3	+	TTCTATAGCCACTGGAATCT	TTTC
	4	+	CCCCAAGCCAGCCCGAACAC	TTTA
zDHHC17	**1**	+	ACACCAAGATGGCGGACGGC	TTTA
	2	**-**	CTTTGTCCGGTTGCCGTACA	TTTT
	3	**-**	CCAAATCAAGCGCTGATTCG	TTTG
	4	+	ATACGAAAACCGGTGAGGTC	TTTG
zDHHC18	**1**	+	ACTGTCCCTACCTGGCTCGC	TTTG
	2	**-**	TCCAGGGCGGCTGCTTCACA	TTTC
	3	+	ACACGTACCTCGTCGCCTCC	TTTC
	4	+	TTCTCTCCCTCTCATTCCTG	TTTA
zDHHC19	**1**	+	CTGCCTTCAATGTGGTGCTG	TTTG
	2	+	CTGTTATCACAGGCTCCCTC	TTTC
	3	+	ACTCAACTTCTCAGACCCTG	TTTC
	4	**-**	GACACCATTGCAGGCGGAAG	TTTG
zDHHC20	**1**	+	TTATGTTTGTATGGTCCTAT	TTTG
	2	**-**	ATAACGTTCCTTTTCAGAAT	TTTC
	3	+	CCTATCTATACCACATCAGC	TTTA
	4	+	CGTGGCTGCAACAGTTTTAG	TTTT

zDHHC21	**1**	+	TTGTTGACCCACATGGTTGG	TTTG
	2	+	CTCACTATGAAGAAGGACAT	TTTC
	3	-	TCCATGTGGGATCTTGGGGT	TTTC
	4	+	ATGAGACCAAAGCGTTCCCA	TTTG
zDHHC22	**1**	+	TTCACCGGCAACTGCATCGG	TTTC
	2	+	CATCTCCTTCGCCCACCCCT	TTTC
	3	-	AGAACCCAGGACAGCTCCGG	TTTC
	4	-	CGAAGACCTCTTGTAAGTTC	TTTC
zDHHC23	**1**	-	TTCACAGGCTTCATACTGCC	TTTC
	2	-	CATGTAATCCATCGATCACA	TTTA
	3	+	TGATCGCCTCCGAATTCCTT	TTTC
	4	-	GTGGGGCTTCCAGCTGGCAT	TTTG
zDHHC24	**1**	-	GCATTGGTAGCAGTAAGCCC	TTTG
	2	+	CCTTGGCCTTCGTGACGGAC	TTTG

Targeted Control sgRNAs

Gene	sgRNA #	Strand	Sequence	PAM
hEIF3B	**1**	-	CTCCCGCGTCCTCGGACACC	TTTC
	2	-	AGTTTCTCAAGTCGGTCGGG	TTTG
	3	+	CCTGGAGTACGCGTCCCCTG	TTTT
hDICER1	**1**	+	GACTGCCATGGCAACAAGAA	TTTG
	2	-	TCGATAGGGGTGGTCTAGGA	TTTC
	3	-	ATAAAGCCCACTTCTGTCAG	TTTC
hPOLR2A	**1**	-	GCAGCGGCCAGTCCGCTCAA	TTTG
	2	-	GGTTGTTAGAGTCCACAAGC	TTTG
	3	-	CTTTGGTCAGATCCTCGTCA	TTTT
hNRF1	**1**	-	GGTCACTCCGTGTTCCTCCA	TTTG
	2	+	CTTCGGAAACTTCGAGCCAC	TTTG
	3	-	ACTATGGTCCGTAGTGCCTG	TTTA

Non-targeted Control sgRNAs

#	0	10	20	30	40	50
0		CAAGTTG GACCGTT AGCTAT	GAATCCC GACGACT TGTTTG	GCGCGGG CCAGACG GCCTTT	GCCGGAA TTGCGCCT CTCGT	GGAACCG GCGATAC CGAACG
01	TAAGGCG CACACCT CGGATG	TCCCGTA GTGGGAT CGCTCC	CGATTATC CTCCACCC GAGC	CAGGCCT TATCGGG ACTCGA	GGTATCG AAATACC GTAGGC	GCGGCCG CGACGTA TACCGT
02	AGGCCGG GCACTGT AAGTTG	AGCGGCG AGATACT TGCCGG	GCCGCGG CTAGGGA CGGACA	ACCGAGG CGCCCCA CAGTGC	ACGGGGG ACGGTCA CACGCG	TACCAAA TCAGAGC GCGGAC
03	TCGGTCA GGCCGTA CAGGTG	CGAGTTG GCTGTACT AGACG	CACCGAG ATCATAC GGAAAA	CCGCCCC CGGAGGT AATAGT	ACACCTCT ACGTGAC ATAGT	
04	TTCGTCCG CTCAGAG TCGAC	GGCGTAA CTCGAGA GTGTGT	GGCCATC GATCCCG GCTTAG	GCGTCGA CTAATGT GGAACA	GCCAACC TGGCATG GTTGAC	
05	AGCAAGG CCTATCCG GCCGT	AGTGCGG CCAAGGG TATTAC	CCGTTGAC GCATCAC CGAGT	AGAACCT AGCGGCT CTCACG	GCCGCAA ATCCTTGT CGGCG	
06	GCCCAAG TTAGCGC ATAGAG	ATGCGTA CGAGGTC TATGAT	CCCCGCGT CGGTCAG GCCTC	ACTCGAT CCACCGT TAACTA	GCTTGTGA CCTGCGCC ATAG	
07	GTCAGCG GTTATTAC CGCAA	GGCGGCG TCCGGCTT GAGTC	ACATTGG GACGGGT TAAACT	TCGAACG CCACCGG TCCCTT	TTACTCGC GGCTCGG ACTCG	
08	GGGCGCC GGGGCGT CCGCGA	GACCGTG TCGGAGC TCGGCC	TCTCGATC GCTAACG TCCCC	TCAGGGG CGACTAC CGTTGT	GCGTGCC AACCATA GAAGCC	
09	TACGATA GGCCGCG GTCTGC	GCCAAGC GATTTTAC GGTTA	TACTAGG CCAGCGA GACTAA	TTGCCGTA TCAGCGA GGAAG	GCAGCAC GGACGAA ACACGG	